

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2001, 11:45:25 Search time 25.1 seconds
(without alignments)
1959.548 Million cell updates/sec

Title: US-09-590-991-6
Port of: 3454

Sequence: 1 YPVVLADTSSSEDALNISDK.....PGKNIKSTLNVINGKSTYGY 664

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 522463 seqs, 74073290 residues
total number of hits satisfying chosen parameters: 5224633

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

Database :

1:	/SID8/gcgdata/genseq/genseq/AAL1960.DAT
2:	/SID8/gcgdata/genseq/genseq/AAL1961.DAT
3:	/SID8/gcgdata/genseq/genseq/AAL1982.DAT
4:	/SID8/gcgdata/genseq/genseq/AAL1983.DAT
5:	/SID8/gcgdata/genseq/genseq/AAL1984.DAT
6:	/SID8/gcgdata/genseq/genseq/AAL1985.DAT
7:	/SID8/gcgdata/genseq/genseq/AAL1986.DAT
8:	/SID8/gcgdata/genseq/genseq/AAL1987.DAT
9:	/SID8/gcgdata/genseq/genseq/AAL1988.DAT
10:	/SID8/gcgdata/genseq/genseq/AAL1989.DAT
11:	/SID8/gcgdata/genseq/genseq/AAL1990.DAT
12:	/SID8/gcgdata/genseq/genseq/AAL1991.DAT
13:	/SID8/gcgdata/genseq/genseq/AAL1992.DAT
14:	/SID8/gcgdata/genseq/genseq/AAL1993.DAT
15:	/SID8/gcgdata/genseq/genseq/AAL1994.DAT
16:	/SID8/gcgdata/genseq/genseq/AAL1995.DAT
17:	/SID8/gcgdata/genseq/genseq/AAL1996.DAT
18:	/SID8/gcgdata/genseq/genseq/AAL1997.DAT
19:	/SID8/gcgdata/genseq/genseq/AAL1998.DAT
20:	/SID8/gcgdata/genseq/genseq/AAL1999.DAT
21:	/SID8/gcgdata/genseq/genseq/AAL2000.DAT
22:	/SID8/gcgdata/genseq/genseq/AAL2001.DAT

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3454	100.0	664	22	AAB4B342	S. pneumoniae Sp12
2	3454	100.0	2120	21	AAV81710	Streptococcus pneu
3	3269	94.6	1007	19	AAW61246	Streptococcus pneu
4	803	23.2	192	19	AAV86159	S. pneumoniae deri
5	453	13.1	1946	19	AAW472273	Lactobacillus bulg
6	447.5	13.0	1962	12	AAW46259	Mutant protease (A
7	440.5	12.8	1962	12	AAV10557	Mutant protease (A
8	438.5	12.7	1962	12	AAV10558	Mutant protease (A
9	435.5	12.6	1968	12	AAV10941	Mutant protease (A
10	435	12.6	1959	12	AAV10562	Mutant protease (A
11	434.5	12.6	1962	12	AAV10560	Mutant protease (A

12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
434.5	434.5	431.5	429.5	343.5	302	302	298	296.5	293	293	292	256	240.5	237.5	237.5	236	236	229.5	229.5	214.5	204	201.5	201	199.5	196	194	194	192	192	188.5	187.5	187.5	
12.6	12.6	12.5	12.4	9.9	8.7	8.7	8.6	8.6	8.5	8.5	8.5	7.4	7.0	6.9	6.9	6.9	6.8	6.8	6.6	6.6	6.2	5.9	5.8	5.8	5.8	5.7	5.6	5.6	5.6	5.6	5.5	5.4	5.4
1962	1962	1974	1560	806	1164	1164	1150	1150	1167	1167	1167	731	690	690	690	654	654	659	659	659	766	892	522	412	382	1316	1337	749	768	394	381	829	21
AAr10561	AAr10563	AAr10940	AAr9415	AAr27481	AAr22469	AAr1263	AAr22471	AArB1265	AArB22470	AArB1264	AArB1266	AArB0822	AArB4904	AAr16337	AArB1787	AArB24129	AArB49841	AArB4121	AArB4121	AArB4123	AArB3304	AArB6810	AArB24132	AArB4838	AArB4836	AArB00246	AArB30505	AArB30504	AArB28897	AArB28896	AArB4261	AArB70053	AArB3302
Mutant protease (N	Mutant protease (K	Mutant protease (d	S. cremoris proteol	Rp-111 residual pr	Streptococcal C4a	Streptococcal C4a	Streptococcal C4a	Streptococcal C4a	Streptococcal C4a	SCP12 peptidase (SCP12 peptidase (Melon protease from	Residual protease-	Bacillus subtilis	Residual protease	Pyrococcus furiosus	Thermophilic stable	Thermophilic stable	MO9656926 Seq. ID 1	Protease. Synthet	Subtilase (St_P2Sc	MprA protein. Bac	Pyrococcus furiosu	Hyperthermostable	Hyperthermostable	Subtilisin N62D/Y11	A calcium-dependen	A calcium-dependen	Arabidopsis thailia	Arabidopsis thailia	Bacillus carlsberg	Bacillus subtilis	Bacillus (Pigeosp)

PT blood infections -
XX
PS Claim 4; Page 47-50; 54pp: English.

CC The invention relates to novel immunogenic polypeptides, Spi28 and Spi30
CC from *S. pneumoniae*. Vaccines comprising the polypeptides are useful for
CC the treatment and prevention of pneumococcal infections, particularly
CC infections caused by *Streptococcus*, such as otitis media, nasopharyngeal,
CC bronchial, lung or blood infections. The antigens are used as immunogenic
CC agents to stimulate an immune response. The antisera and antibodies may
CC also be used in diagnosing and treating pneumococcal infections.
CC Recombinant polypeptides serve as a mechanism for stimulating production
CC of antibodies for use in passive immunotherapy, diagnostic reagents, and
CC as reagents in other processes such as affinity chromatography. The
CC present sequence represents the *S. pneumoniae* Spi28 polypeptide.

XX
XX Sequence 664 AA:

Query Match 100.0%; Score 3454; DB 22; Length 664;
Best Local Similarity 100.0%; Pred. No. 6 Be-195;
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YPVVLADTSSSEDALNISDEKVAENKEHENIHSAMETSDQFEKKTAVIKEEVYSKN 60
DB 1 YPVVLADTSSSEDALNISDEKVAENKEHENIHSAMETSDQFEKKTAVIKEEVYSKN 60
QY 61 PYIDNNTSNEAKIKKEENSNKSOGDYTDSPVKNKTEPNKKEKDYVYIAEFKDSKESGKAI 120
DB 61 PYIDNNTSNEAKIKKEENSNKSOGDYTDSPVKNKTEPNKKEKDYVYIAEFKDSKESGKAI 120
QY 121 KELSLKNTKVLVYTDRIFNCSAIEETPDNLKIKOIEGISSVERAOKVQPMNHARKEI 180
DB 121 KELSLKNTKVLVYTDRIFNCSAIEETPDNLKIKOIEGISSVERAOKVQPMNHARKEI 180
QY 181 GVEBAIDYLSINAPFCKNEDGRCGVISNIDTGYRRHKAIRIDDAKASRREKEDIKG 240
DB 181 GVEBAIDYLSINAPFCKNEDGRCGVISNIDTGYRRHKAIRIDDAKASRREKEDIKG 240
QY 241 TDKRWYISDKRPHAFNYNGKIVKVEYDGRDYFDPHGMIAIGILGNDPEDIKKNNG 300
DB 241 TDKRWYISDKRPHAFNYNGKIVKVEYDGRDYFDPHGMIAIGILGNDPEDIKKNNG 300
QY 301 IDGIAPNAOIFSYKMSDAGSGFAGDETFMFAIDSIKHNVDVYSSSGFTGTGLVGERKY 360
DB 301 IDGIAPNAOIFSYKMSDAGSGFAGDETFMFAIDSIKHNVDVYSSSGFTGTGLVGERKY 360
QY 361 WQAIRALRKAGIPMYVATGNATASASSSSMDLVANHLKMTDTGNTVTTAHEDAIYAS 420
DB 361 WQAIRALRKAGIPMYVATGNATASASSSSMDLVANHLKMTDTGNTVTTAHEDAIYAS 420
QY 421 AKKQTVFEDKYNIGSEFFKRNIGAFDKSKITTNEDGTRKPKLKFFYIKGDDODLIG 480
DB 421 AKKQTVFEDKYNIGSEFFKRNIGAFDKSKITTNEDGTRKPKLKFFYIKGDDODLIG 480
QY 481 LDLRGKIAVMDRIYTKDLKNAFKKAMDGARAIVNVTYNNRDNMTLPRAMGYEADG 540
DB 481 LDLRGKIAVMDRIYTKDLKNAFKKAMDGARAIVNVTYNNRDNMTLPRAMGYEADG 540
QY 541 TKSQVFEISGDDGKLMNMNINPKKTEVKRNKEDFKDLKQYPIPMESNSKNKPNVG 600
DB 541 TKSQVFEISGDDGKLMNMNINPKKTEVKRNKEDFKDLKQYPIPMESNSKNKPNVG 600
QY 601 EKEIDFAPDTEKELKEDIIVPAGSTSMGPRIDLKLPVSAFGKNIKSTLVNTYINKS 660
DB 601 EKEIDFAPDTEKELKEDIIVPAGSTSMGPRIDLKLPVSAFGKNIKSTLVNTYINKS 660
QY 661 TYGY 664
DB 661 TYGY 664

RESULT 2

AAV81710
ID AAV81710 standard; Protein; 2120 AA.

XX
XX AAV81710;

DT 02-JUN-2000 (first entry)

XX Streptococcus pneumoniae protein sequence ID3.

XX Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;
KW pneumococcal septicemia; sinusitis; meningitis; therapy.

XX Streptococcus pneumoniae.

XX WO200006738-A2.

XX 10-FEB-2000.

PD 27-JUL-1999; 99WO-GB02452.

PR 27-JUL-1998; 98GB-0016336.

PR 19-MAR-1999; 99US-0125329.

PA (MICR-) MICROBIAL TECHNICS LTD.

PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;

DR WPI; 2000-195301/17.

DR N-PSDB; AA291806.

PT Streptococcal proteins and polynucleotides useful for diagnosis,

PT treatment and prophylaxis of bacterial infections -

XX Claim 2; Page 41-42; 76pp: English.

XX This sequence represents a *Streptococcus pneumoniae* protein of the
XX invention. The proteins (or their homologues, derivatives and/or
XX fragments) are useful as immunogens or antigens. Immunogenic or antigenic
XX compositions comprising the proteins are useful as vaccines and also in
XX diagnostic assays. The sequences are useful for the detection or
XX diagnosis of *S. pneumoniae* infection, by contacting a sample to be tested
XX with them. Agents capable of antagonising, inhibiting or interfering with
XX the function or expression of the protein or polypeptide are useful in
XX medical compositions in the treatment or prophylaxis of *S. pneumoniae*
XX infection. As the sequences can be used to treat *S. pneumoniae* infection,
XX they can be used to treat bacterial pneumonia, which has high rates in
XX young children, the elderly, and in patients with predisposing conditions
XX such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
XX or with immunosuppressive disorders, especially AIDS. They can also be
XX used to treat pneumococcal septicemia, otitis media, sinusitis, and
XX meningitis.

XX Sequence 2120 AA:

Query Match 100.0%; Score 3454; DB 21; Length 2120;
Best Local Similarity 100.0%; Pred. No. 2,9e-194;
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YPVVLADTSSSEDALNISDEKVAENKEHENIHSAMETSDQFEKKTAVIKEEVYSKN 60
DB 1 YPVVLADTSSSEDALNISDEKVAENKEHENIHSAMETSDQFEKKTAVIKEEVYSKN 60
QY 61 PYIDNNTSNEAKIKKEENSNKSOGDYTDSPVKNKTEPNKKEKDYVYIAEFKDSKESGKAI 120
DB 61 PYIDNNTSNEAKIKKEENSNKSOGDYTDSPVKNKTEPNKKEKDYVYIAEFKDSKESGKAI 120
QY 121 KELSLKNTKVLVYTDRIFNCSAIEETPDNLKIKOIEGISSVERAOKVQPMNHARKEI 180
DB 121 KELSLKNTKVLVYTDRIFNCSAIEETPDNLKIKOIEGISSVERAOKVQPMNHARKEI 180

```

QY 181 GVEAIDYKSTNAPPGKNGFNGNYSINIDTCTDYRHKRAMIDDDAKASMRKKEDLKG 240
    |||||||
DB 181 gveaaidyKstnappgkngfngnysinidctdyrHkramiddakasmrkkedlkg 240
QY 241 TDKNYLSKIPAFNYVNGSKITVEKYDDGRYFDPHGHTAGIAGNDETDIKNPNFNG 300
    |||||||
DB 241 tdKNyLSkIPaFnyVngSkITveKyDDGRyFDPHGHTAGIAGNDETDIKnPNfNG 300
QY 301 IDGIAPNAOIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLVGEKY 360
    |||||||
DB 301 idgiapnaoifSyKmySDaGSgFAGdeTFmFAIEdSIkHNvDVSVSSgFTgtGLvGEKy 360
QY 361 WQAIRLRKAGIPMYVATGNVATSSASSSDLVANNHLMKMTDGTNVRTIAHEDAIIVAS 420
    |||||||
DB 361 wQaiRLrKAGiPMvVatGnVATSSASSsDLVANNhLMkMTDGTnVRTIAhEDAIIVaS 420
QY 421 AKKQTEPFPKVNITGSESPFRNIGAFEDSKITNEDGTRKAPSKLFEVYIGKODDGLIG 480
    |||||||
DB 421 aKqTeFpKvNiTgSeSPFRniGAFEdSKITnEDGtRKaPSKLfEvYiGKOdDGLiG 480
QY 481 LDRGRIAVMDRIYTTDLKNAFKKAMDKGARAIMVTVNVYVRDMMWTELPAMGYEADG 540
    |||||||
DB 481 lDrGRiAVMDriYtTDLkNAfKKAMdKGARaiMVTVNVyVRdMMWTElPaMGyEAdG 540
QY 541 TKSQVPSISGDDGVKLMNMNINPDKTEVRKRNKEDEFKDLEOYYPIDMESFNSNKNPVGD 600
    |||||||
DB 541 tKsQvPSiSGdDgVklMnmNiNpDKtEvRkRNkEdEfKdLEoyYPIDMeSfNSnKpVgD 600
QY 601 EKRIDKFPADPTKELKEDITYPAGSTSGRPRIIDLLKRPDVAPSKNIKSTLVNNGKS 660
    |||||||
DB 601 eKriDkFpAdPtKELkEdITyPaGStSGRPRIIDLLkRPdVAPSKniKStLVnNGKs 660
QY 661 TTYG 664
    ||||
DB 661 tYgY 664

```

RESULT 3
 AAM61246
 ID AAM61246 standard; Protein; 1007 AA.
 XX
 AC AAM61246;
 XX
 DT 02-OCT-1998 (first entry)
 XX
 DE Streptococcus pneumoniae SPI22 protein.
 XX
 OS Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
 XX detection; pneumonia; otitis media; meningitis.
 XX
 PN MO9818930-A2.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97MO-US19422.
 XX
 PR 31-OCT-1996; 96US-0029960.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Chol GH, Hromockyj A, Johnson LS, Kunsch CA;
 XX
 DR MPI: 1998-272224/24.
 XX N-PSDB; AAV27431.
 XX
 PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
 XX pneumoniae - or their epitope-containing fragments, useful in
 XX protective or therapeutic vaccines, and for diagnosis
 PS Claim 11; Page 92-93; 118pp; English.
 XX

CC The present sequence represents a protein from Streptococcus pneumoniae.
 CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
 CC can be useful in vaccines for inducing protective antibodies against
 CC Streptococcus pneumoniae. For treatment or prevention of infection e.g.
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 CC are used to detect Streptococcus infection (by usual hybridisation or
 CC amplification methods), also for isolating Streptococcus genes or their
 CC allelic variants. The protein can be used similarly to detect specific
 CC antibodies in standard immunoassays, especially for diagnosing or
 CC monitoring infections. Antibodies which bind the protein are used to
 CC detect corresponding antigens, to purify the protein and for passive
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000
 CC (especially 10-300) mu g/ml per dose.
 XX
 XX
 SO Sequence 1007 AA;

Query Match 94.6%; Score 3269; DB 19; Length 1007;
 Best Local Similarity 100.0%; Pred. No. 8.2e-184;
 Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 38 ETSODERREKKTAVIKREYVSKNPVIDNNTSNEAKIKENSNSKSGDYYDTSFVNKNTEN 97
    |||||||
DB 1 etsqdEkkKtAvIkReYvSkNpViDnNtSNeAkIkEnSnsKsGdYyDtsfVnKnTEn 60
QY 98 PKKEDRVVYIAEERKDESGEKAKEISLKNKTVLYTYDRIFFNGSAIETTPDMXIKOI 157
    |||||||
DB 61 pKkedrvVyIaeERkDESGeKaKEiSLkNkTVLyTYdRIffNGsAIETTPDMXiKoI 120
QY 158 EGISVSERAKQVOPMMNHARKEIGVEAIDYLSINAPPGKNGFGNWSINIDGTDYR 217
    |||||||
DB 121 egIsVsErAkQvOPmMnHARKEiGvEAIdYlSiNaPPgKngFGnWsiNiDgTDyR 180
QY 218 HKRAMRIDDDAKASMRKKEDLKTOKNMYLSDKIRIPAFNYNGSKITVEKYDDGRYRPP 277
    |||||||
DB 181 hKraMriDDaKASmrKKEdLkTOKnMYlSDKiRiPaFnyNGSkITveKyDDGRyRPP 240
QY 278 HGMHIGILAGNDTODIKNFNGIDGIAIPNAQIFSYKMYSDAGSGFAGDETFMFAIEDSI 337
    |||||||
DB 241 hgmHigIlAGndtODiKnfNGIDgiAPnaQIFsYkMySDaGSgFAGdeTFmFAIEdSI 300
QY 338 KHNVDVSVSSGFTGTGLVGEKYWQAIRLRKAGIPMYVATGNVATSSASSSDLVANNH 397
    |||||||
DB 301 kHNvDVSVSSgFTgtGLvGEKyWQaiRLrKAGiPMvVatGnVATSSASSsDLVANNh 360
QY 398 LKMTDGTNVRTIAHEDAIIVASAKQTVFDFKVNITGSESPFRNIGAFEDSKITNED 457
    |||||||
DB 361 lKmtDGTnVRTIAhEDAIIVaSaKqTVfDFkVNiTgSeSPFRniGAFEdSKITnEd 420
QY 458 GTRKAPSKLFEVYIGKODDGLIGLDRGIAVMDRIYTTDLKNAFKKAMDKGARAIMVTV 517
    |||||||
DB 421 gTrKaPSKLfEvYiGKOdDGLiGLDrGiAVMDriYtTDLkNAfKKAMdKGARaiMVTV 480
QY 518 TVNVYVRDMMWTELPAMGYEADGTRKQOVESISGDDGVKLMNMNINPDKTEVRKRNKEDEK 577
    |||||||
DB 481 tvnYvVRdMMWTElPaMGyEAdGtRKQovESiSGdDgVklMnmNiNpDKtEvRkRNkEdE 540
QY 578 DKLEOYYPIDMESFNSNKNPVNGDEKEDFKFAPDPTKELKEDITYPAGSTSGRPRIIDLL 637
    |||||||
DB 541 dKLEoyYPIDMeSfNSnKpVNgDEkEdFKfAPDPTkELkEdITyPaGStSGRPRIIDlI 600
QY 638 LKPDYSAFGKNIKSTLVNNGSKTYGY 664
    |||||||
DB 601 lKpdYsaPgKnIkStLVnNGskTygY 627

```

RESULT 4
 AAY86159
 ID AAY86159 standard; Protein; 192 AA.
 XX
 AC AAY86159;
 XX

Qy	408	RTAAHEADIAVASAKNOTVEEDKV--NIGSEPKRYRN--TGAFEDSKITTTNEDGTka--	466
Db	412	-----pdallvsaemskvtldvckdeIlgvltfssnslkga---aqvltqlsnysvl	462
Qy	462	PSKLFYVYIGKGQODDlG---IDLRGKlAVMDR-IYTKDLKNAFKKADMGARAlMYVN	517
Db	463	tkklklkldmglyggaddylaekkeavkqglavkrqytlfsakvanaka--agaaglvlyN	520
Qy	518	TVNRYNRDNTLPMAGYEADegTKSOVSISGDDGvKLMNMINPDKKTEVKRNKKEDPK	577
Db	521	-----seddg--llsmajldcktfrcLjgmskadqk	547
Qy	578	DKLEQYVPIDMESFNNSKNPNVGD---KEIDFKFAPDTOKELKEDlIYPAGSTSMKPRl	634
Db	548	fwlqyqkvkvaasrlkfgtaltalidnsragkmsdf-----tswgpvc	586
Qy	635	DLLKPDVSAFGKNIKSTLN	654
	587	eldfxfpeltapqgklyalan	606
RESULT	6		
ID	AAR10559		
XX	AAR10559	standard; Protein; 1962 AA.	
XX	AG	AAR10559;	
XX	DT	15-Apr-1991 (first entry)	
XX	DE	Mutant protease (A137G/K138L/T139A).	
XX	KW	Mutant protease gene; fermentation; foodstuff; flavouring;	
XX	KM	lactic acid bacteria.	
OS		Lactococcus lactis SK11.	
XX	Key	Location/Qualifiers	
FX	Peptide	1..187	
FT	Protein	/label= signal_peptide	
FT		188..1962	
FT		/label= mature_protein	
XX	EP411715-A.		
XX	06-FEB-1991.		
XX	02-AUG-1990;	90EP-0202113.	
XX	04-AUG-1989;	89NL-0002010.	
XX	(NEZU-) NED INST ZUIVELONDE.		
XX	Vos PAJ, Slezén RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;		
XX	WPI; 1991-038622/06.		
XX	N-PSDB; AAQ10413.		
XX	Mutant protease gene(s) and protease(s) - derived from type I and		
XX	III protease genes from Lactococcal strains, used in fermentation		
XX	foodstuffs and flavourings		
XX	Disclosure: Fig 1(1-7)+5(b); 29 pp; English.		
XX	The mutant protease having new cleavage specificities is		
XX	obtained by replacing three amino acids.		
XX	This mutant may then be used to prepare hybrid proteases,		
XX	the fusion being between a type I and a type III protease of		
XX	L.lactis Wg2 and SK11.		
XX	The product has modified properties, e.g. thermostability,		
XX	alkaline/acid pH stability, oxidative stability, autoprocolysis etc.		
XX	compared to the parent protease(s). The proteases can be used for		
XX	preparing products (butter cheese, human and animal foodstuffs)		

CC	prepared with the aid of lactic acid bacteria.	
CC	See also AAQ10411-17 and AAQ10870-71.	
XX	Sequence 1962 AA:	
50		
Query Match	13.0%; Score 447.5; DB 12; Length 1962;	
Best Local Similarity	27.3%; Pred. No. 6.3e-18;	
Matches 166; Conservative	92; Mismatches 236; Indels 111; Gaps	25
QY	78 NSNKGCDYTSFVKNKFNENPKEDKV--YIAEFKDESEKAIKELSLKNTFVLYT	134
DB	84 nhvqgqddlyvdvlyvmsaa-pasenglllrltdysstcclqecnkvlaaqasvkaaveqvt	142
QY	135 -----YDRIFNGSAIETTPDNLDKIKOEGISSVERAOKVOPMNNARKKEIGVEAID	187
DB	143 qqlagesygyuvvfnstckvrvvdipklkqlaagvktclakvyvplcakansmanvqavvs	202
QY	188 YLKSINAFPGKNFDRGNVINSIDTGTGYRNKAMRIDDAKSMRKKEDL-KGTP---K	243
DB	203 nyk-----ykhgegtvsvidsagldprthdmtlsdd-kvklksdveftclvkh	251
QY	244 NYWLSDKIPHAFFNY-NGSKITVERKYDDORDYFDPHGMHIAIGLGNDEODIKENGID	302
DB	252 gryfnskrrpygfnygdndlttdckvd-----eqhgmhvaglligangydd--paksyv	303
QY	303 GIAPNAQIFSKMYSDAG-SCFAGDETFMHAIEDSIKHNVDVSVSSGF-TGTGLVGEKY	360
DB	304 gvapeaqqlamkvfnstsdtslgsatlvsvaledsaklgadvlnmlsgnsgqtledd	363
QY	361 WQATFALKKAGIPNVVATGNVATSSSSSMDLVANNHLKITDTGNTRTAAHEDALAVAS	420
DB	364 laayvnaaesglvaaviasgnsgtsat--egvlnkdyyuylqdnemvgsrptargatvas	421
QY	421 AKNGVFEEDKVINI-----GGESEFKY--RMIGAFFDSKITTTNEDGTAKPSKLKFFVY	470
DB	422 aentlvltqavtlctgtqlqigpeltqlqshnfctgsfdokkfyyivkdaag-----nl	473
QY	471 GKGODDILIGDLRGKIAVMDR-IYTKDLKNAFKKAMDGARAINVNTVNYNRDNMTE	529
DB	474 skgaladyta-dakgkialvkrgefsfdokklyaqe--agaaglllvnt-----dgtc	523
QY	530 LPAKMYEADDEGKSOVFSISGDDCYKL-WNNINPDKTEFVKNRKNKEDKDKLEQYPIID	587
DB	524 lpmtsialt-llftpfqlsvtgklydvwvleahpddaslgvk-----	563
QY	588 MESFNSNKPWNGDEKEDKFAFDMDKELKEDDIYPAQSTSGMPRIDLLKPDVAPAK	647
DB	564 -----lltlaipn---qkytcdkm--sdctsygprvsnlsfkpdlcpagp	602
QY	648 NIKSTLN 654	
DB	603 nlwstgn 609	
RESULT 7		
AAQ10537		
ID	AAQ10537 standard; Protein; 1962 AA.	
XX	AAQ10537;	
AC	15-APR-1991 (first entry)	
XX		
DE	Mutant protease (A137G/K138D).	
XX		
KW	Mutant protease gene; fermentation; foodstuff; flavouring;	
XX		
OS	Lactococcus lactis SK11.	
XX		
EH	Key	Location/Qualifiers
FT	Peptide	1..187
FT		/label=signal_peptide

Best Local Similarity 27.2%; Pred. No. 2.1e-17;
Matches 165; Conservative 92; Mismatches 239; Indels 111; Gaps 25;

```

QY 78 NSKSGODTDSVNNKTEPRKEDKV---YIAEFKDSGSKAIKELSLNTKVLTY 134
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 84 nkvgqddlyvdvlgmsaa-paseengllrtcdysaeiqetnkvlaagvaaveqvt 142
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 135 -----YDRIFNGSAIETTPDLNDIKIQEGISSVERAOKVOPMNNHARKEIGVEAID 187
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 143 qqtagesygyvvnvgfctkvrvvdiprkqkqagkytvtlakyvrrptcdakamsnvaqaws 202
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 188 YLKSINAPFGKNFDGKGMVSNIDTGTDRHKKAMRIDDAKASMRKKEDL-KGTD--K 243
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 203 nyk-----ykgegtvsvsvidsgldprchkdmrlsdd--kdvkklksdvекftdlvkh 251
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 244 NYWLSOKIIPAHFNYY-NGSKITVEKYDDGRDYDFDPHGMHAGIAGMDTODIKNFPGID 302
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 252 gryfnskvpygfnvadndtlldckvd-----eqgmhvaagllgngtgdd--paksyv 303
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 303 GIAPNAQIESYKMYSDAG-SGFAGDETMFHAIEDSIKHNVDVSVSSGF-TGTGLVGEKY 360
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 304 gvaapeagllamkvfnstdtsqpsatvsaiedsaklgadvlnmslgsnsgnqtlledpe 363
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 361 WQAIRALKRAGIPMVVATGNYATSSASSWDLVANNHLKMTDTGNVTRTAHEDALAVAS 420
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 364 laavqnaesgtaavlaagsnsgtsaat--egvnykdyglqdnemvgspsgrgattvas 421
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 421 AKKQTFEFDVNI-----GGESFKY--RNIGAFPKDKITTNEDGCTAPSKIKVYI 470
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 422 aentldvltgavtltdgtqlgipetqlshdltgsdqkkylylvdaag-----nl 473
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 471 GKGODDGLGLDGRKIAVMDR-IYTRDLKNAFKKADKARAIMVNTYNNYRDNMTE 529
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 474 skgaladyta-dakgkialavkrgefsfdckkyaga--agaaglllynt-----dgt 523
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 530 LPMAGTEADEGTSQVFSISGDDGVKL--WNMINPDKTEVKRNKNEDEKLEQYYPID 587
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 524 tpmstalt--tftptglsvsgtklvdvtahpddslgvk----- 563
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 588 MESFNSKKRPVSGDEKEIDFPAPDTEKLYKEDLIYPAGSTSGRPRIIDLKPDVSAKG 647
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 564 -----ltlamlpn--qkytedkm--sdltsygsprvsnlsfkpdlitap9 602
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 648 NIKSTLN 654
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 603 nlwstqn 609
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      ULT 9
      ID AAR10941 standard; Protein; 1968 AA.
      AC AAR10941:
      DT 15-APR-1991 (first entry)
      DE Mutant protease (delta137-139/Ins9).
      KM Mutant protease gene; fermentation; foodstuff; flavouring;
      KM lactic acid bacteria.
      XX Lactococcus lactis SK11.
      XX OS
      XX Key Location/Qualifiers
      FH Peptide 1..187
      FT /label= sig_peptide
      FT 188..1968
      FT Protein /label= mat_protein
      XX BP411715-A.
      XX PN
      XX PD 06-FEB-1991.
      XX

```

```

PF 02-AUG-1990; 90BP-0202113.
XX
PR 04-AUG-1989; 89NL-0002010.
XX
PA (NEU-) NED INST ZUIVELONDE.
XX
PI Vos PAJ, Slezien RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
DR WPI: 1991-038622/06.
XX N-PSDB; AAQ10871.
XX
PT Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
PS Disclosure; Fig 1(1-7)+5(b); 29pp; English.
XX
CC The wild-type L.lactis SK11 protease gene sequence was determined
CC by the applicant (EP-307011).
CC The mutant protease having new cleavage specificities is
CC obtained by deleting three amino acids (line bps) and inserting
CC 9 other residues.
CC This mutant may then be used to prepare hybrid proteases,
CC the fusion being between a type I and a type III protease of
CC L.lactis Wg2 and SK11.
CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC compared to the parent proteases(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
CC See also AAQ10411-17 and AAQ10870-71.
XX
SQ Sequence 1968 AA:
      Query Match 12.6%; Score 435.5; DB 12; Length 1968;
      Best Local Similarity 27.1%; Pred. No. 3.2e-17;
      Matches 166; Conservative 92; Mismatches 238; Indels 117; Gaps 26;
      QY 78 NSKSGODTDSVNNKTEPRKEDKV---YIAEFKDSGSKAIKELSLNTKVLTY 134
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      DB 84 nkvgqddlyvdvlgmsaa-paseengllrtcdysaeiqetnkvlaagvaaveqvt 142
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      QY 135 -----YDRIFNGSAIETTPDLNDIKIQEGISSVERAOKVOPMNNHARKEIGVEAID 187
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      DB 143 qqtagesygyvvnvgfctkvrvvdiprkqkqagkytvtlakyvrrptcdakamsnvaqaws 202
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      QY 188 YLKSINAPFGKNFDGKGMVSNIDTGTDRHKKAMRIDDAKASMRKKEDL-KGTD--K 243
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      DB 203 nyk-----ykgegtvsvsvidsgldprchkdmrlsdd--kdvkklksdvекftdlvkh 251
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      QY 244 NYWLSOKIIPAHFNYY-NGSKITVEKYDDGRDYDFDPHGMHAGIAGMDTODIKNFPGID 302
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      DB 252 gryfnskvpygfnvadndtlldckvd-----eqgmhvaagllgngtgdd--paksyv 303
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      QY 303 GIAPNAQIESYKMYSDAG-SGFAGDETMFHAIEDSIKHNVDVSVSSGF-TGTG 354
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      DB 304 gvaapeagllamkvfnstdtsqpsatvsaiedsaklgadvlnmslgsnsgnq 363
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      QY 355 LGEKYYQAIRALKRAGIPMVVATGNYATSSASSWDLVANNHLKMTDTGNVTRTAHED 414
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      DB 364 tledpelaavqnaesgtaavlaagsnsgtsaat--egvnykdyglqdnemvgspsgrg 421
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      QY 415 AIAVASAKNTEFEDKNI-----GGESFKY--RNIGAFPKDKITTNEDGCTAPSK 464
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      DB 422 atcvaseentdvltgavtltdgtqlgipetqlshdltgsdqkkylylvkaasg---- 477
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      QY 465 LKEFYIGKODDGLGLDGRKIAVMDR-IYTRDLKNAFKKADKARAIMVNTYNNY 523
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      DB 478 -----nlsgaladyta-dakgkialavkrgefsfdckkyaga--agaaglllynt----- 525
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      QY 524 RDMWTELPAMGYEADGTSQVFSISGDDGVKL--WNMINPDKTEVKRNKNEDEKLE 581
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db	526	-dglatpmtsalatl--ttftcflgssavtqkllvdtwltapddslgkx-----	56p
Oy	562	OYPTIDWSEFNSKPNWNGDEKEIDFKRAPDTDKELTKEDIIIVAGSTSWGSPRIDLLIKPD 6411	
Db	570	-----ltlam1pn---qkyredkm---sdfstsyppvsnlsfkipd 6022	
Oy	642	VSAPGKNIKSTPLN 654	
Db	603	ltapgnlwtstqn 615	
	RESULT 10		
ID	AAR10562		
XX	AAR10562 standard; Protein: 1959 AA.		
XX	AAR10562:		
AC			
XX			
DT	15-APR-1991 (first entry)		
XX			
DE	Mutant protease (delta137-139).		
XX			
XX	Mutant protease gene; fermentation; foodstuff; flavouring;		
XX	lactic acid bacteria.		
XX			
OS	Lactococcus lactis SK11.		
XX			
FT	Key	Location/Qualifiers	
FT	Peptide	1..187	
FT	Protein	/label= signal_peptide	
FT		188..1959	
XX		/label= mature_peptide	
PN	EP411715-A.		
XX			
PD	06-FEB-1991.		
XX			
PP	02-AUG-1990; 90EP-0202113.		
PR	04-AUG-1989; 89NL-0002010.		
XX			
PA	(NEZU-) NED INST ZUIVELONDE.		
PI	Vos P.M.J., Slezzen R.J., De Vos W.M., Kok J., Venema G., Haandrikman A.J.		
XX			
DR	WPI: 1991-038622/06.		
XX	N-PSDB: AAQ10416.		
XX			
PT	Mutant protease gene(s) and protease(s) - derived from type I and		
PT	III protease genes from lactococcal strains, used in fermentation		
XX	foodstuffs and flavourings		
PS	Disclosure: Flg 1(1-7)+5(b); 29 pp; English.		
XX			
XX	The mutant protease having new cleavage specificities is		
CC	obtained by deleting three amino acids.		
CC	This mutant may then be used to prepare hybrid proteases,		
CC	the fusion being between a type I and a type III protease of		
CC	<i>L.lactis</i> Wg2 and SK11.		
CC	The product has modified properties, e.g. thermostability,		
CC	alkaline/acid pH stability, oxidative stability, autolysis etc.		
CC	compared to the parent protease(s). The proteases can be used for		
CC	preparing products (butter cheese, human and animal foodstuffs)		
CC	prepared with the aid of lactic acid bacteria.		
CC	See also AAQ10411-17 and AAQ10870-71.		
XX			
XX	Sequence 1959 AA;		
XX			

Query Match	12.6%:	Score 435;	DB 12;	Length 1959;
Best Local Similarity	27.1%:	Pred. No. 3.4e-17;		
Matches 164;	Conservative	92;	Mismatches 238;	Indels. 112;
				Gaps 25;

```

Db      84  nkvgqddlvadlvqmsaa--pasengllrtcdysstaetqetlnkvlaagsvaaveqv 142
Oy      135  -----YDRIFNGSALETTPDNLNDKIOEGISSVERAOKVORPMNHAKETIGVEAID 187
Db      143  qtlagesygyvungfstekvrvvdiprkltqglavutvlelvayvptdaksmanavgaws 202
Oy      168  YLKSINAFEGKNPFGRCGVINISINDTDYRHKARIRIDDKAKSRRFEKDL-KGTD---K 243
Db      203  nyk-----yagegtvsvlsgslorpchkmrlsdd--kdkklksdvekfctdvkh 251
Oy      244  NYWISDKIPAHFNYY--NGKITVERKYDDGRDYEDPHGMIAGTILAGNDTEODIKNFENGID 302
Db      252  gryfsnksyvygfnayadndctlddkvd-----eqhgmhvaglllgangtqgd--paksv 303
Oy      303  GIAPMAOIEFSXKMYSDAGSGFAGBETPMHAIEDSIKHNVPVSVSSGF--GTGTGVGERKYW 361
Db      304  gvapaagqlhankvsn--sdtsgsaetvvasledsaklgadvlmslsgnsnqtlepdl 361
Oy      362  QAIRALKRAGIPMYVATGNVATSSSSSMDLVANNHLKMTDTCNVYTRAAHEDAIVASA 421
Db      362  aavqanageslaavlsagnsgtsgsat--egvnydyugldnemvsgpgrsgetvasa 419
Oy      422  KNQYVEPCKYMI-----GGESEFYK--RNIGAFPKPSKITINEDSTKAPSKLKFYYIG 471
Db      420  entdvlqtaavltcgtqlqjrlpeltqjshnftysfdqkfflyvkdasg-----nls 471
Oy      472  KGODDGLIGDLRGKIAVMDB-IYTKDLKNAFKRAMDKGARALIMVYTVYVYNRDMWTEL 530
Db      472  kgaladyta-dakgkialavkrgefsfdckqyaga--agaaglllvct-----dgtat 521
Oy      531  PAMGEADEGRKSGVESISGDDGK/L--WMNINIDPKKTEYKRNKKEDFKKLEQYVIDM 588
Db      532  pmtstaltl--ltfptfgtssvtygkltvdwvctahpddslayk-----560
Oy      589  ESFNSNKNPVGDEKEDEIKFFKAPDPDKELIYKEDIIVPAGSTSWGPRRIDLLKPPVSAFGN 648
Db      561  -----ltlamipn---kqlytedkm--sdfctsygpyvsnlsfcpdltpagyn 600
Oy      649  IKSTLIN 654
Db      601  lwtstqn 606

```

RESULT	11
AAR10560	
ID	AAR10560 standard; Protein; 1962 AA.
XX	
AC	AAR10560;
XX	
DT	15-APR-1991 (first entry)
XX	
DE	Mutant protease (K138N).
XX	
KW	Mutant protease gene; fermentation; foodstuff; flavouring;
KW	Lactic acid bacteria.
XX	
OS	Lactococcus lactis SK11.
XX	
FH	Key
FT	Peptide
FT	Location/Qualifiers
FT	/label= signal_peptide
FT	Protein
FT	188..1962
FT	/label= mature_protein
XX	
PN	EP411715-A.
XX	
PD	06-FEB-1991.
XX	
PF	02-AUG-1990;
XX	
PR	04-AUG-1989;
XX	
XX	89NL-0002010.
XX	


```

QY 188 YLKSNAPFGKNGVNSINIDTGTDRHKAIRIDDDAKASRFKKEDL-KGTD--K 243
D 203 nyk-----ykgeglvsvlsgldprtkhmlsd--kdklksdvckftdvkh 251
QY 244 NYWLSDKIPAHFNY--NGKLTVEKYDDGRDYFDPHGHNIAGLIGAGNTEODINFGCID 302
D 252 gryfnskvpygfyndndlttdkvd-----eqghmvagllgagngtd--paksyv 303
QY 303 GIAPNAOFESKMYSDAG-SGFADETMFHAIEDSIKHNVDVSVSSGF-TGTGLVGEKY 360
D 304 gvapeagllamkvfsnctsanctsaclvsaiedsaklgadvlmslgsdngnltledpe 363
QY 361 WQAIRALKAGIPMVAVTGNATSASSSMDLVANNHLMKMTDGTGNVTRTAHEDAIIVAS 420
D 364 laavqanesslaavvsgnsgtsat--egvnhkdygldnemvsgpysgtsatlvaa 421
QY 421 AKNOTVEFDKYN-----GGESFNY--RNIGAFPKSKITTEDGTAKPSKLFVYI 470
D 422 aentdvltqavltldgltglqldpeltqlshndfctsgtdqkfyivkdasg-----n 473
QY 471 GKGDODLIGDLRGKIAVMDR-IYTKDLKNAFKKAMDGAIRIMVNTVNYNRDMTE 529
D 474 skgaladyta-dakgkialvkrgefsfdkqkyaga--agaagllivnt-----dgt 523
QY 530 LPAMGYEADGKTSQVFSISGDDGVKL--WNMINDPKKTEVKRNKKEDDFKDLQYYPID 587
D 524 tpmstalt--ctfptfglsvtgkivdwvtahpddslgk----- 563
QY 588 MESFNSKPNVGDKEIDFKAPDPTDKELYKEDIIVPAGTSWGPRIIDLKPPVSAAGK 647
D 564 -----ltlamlpn--qkyledkm--sdfstgypvanlsfkdptltpg 602
QY 648 NIKSTLN 654
D 603 niwstgn 609

RESULT 13
AAR10563
ID AAR10563 standard; Protein; 1962 AA.
AC AAR10563;
XX
XX 15-APR-1991 (first entry)
XX
DE Mutant protease (K748T).
XX
XX Mutant protease gene; fermentation; foodstuff; flavouring;
XX lactic acid bacteria.
XX
XX Lactococcus lactis SK11.
XX
FH Key Location/Qualifiers
FT Peptide 1..187
FT /label= signal_peptide
FT Protein 188..1962
FT /label= mature_protein
XX
XX EP411715-A.
XX
XX 06-FEB-1991.
XX
XX 02-AUG-1990; 90EP-0202113.
XX
XX 04-AUG-1989; 89NL-0002010.
XX
XX (NEZU-) NED INST ZUIVELONDE.
XX
XX Vos PAJ, Slezzen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX WPI: 1991-038622/06.
XX
XX N-PSDB: AAO10417.
XX

```

```

PT Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
XX PS D1sclosure; Fig 1(1-7)+5(a); 29 pp; English.
XX
XX The mutant protease K748T having new cleavage specificities is
CC obtained by carrying out single amino acid substitutions.
CC This mutant may then be used to prepare hybrid proteases,
CC the fusion being between a type I and a type III protease of
CC L.lactis Wg2 and SK11.
CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC compared to the parent protease(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
CC See also AAO10411-17 and AAO10870-71.
XX
XX Sequence 1962 AA:
XX
XX

```

```

Query Match 12.68; Score 434.5; DB 12; Length 1962;
Best Local Similarity 27.08; Pred. No. 3.0e-17;
Matches 164; Conservative 92; Mismatches 240; Indels 111; Gaps 25;

```

```

QY 78 NSNKSQGDYTSFYNNKNTENPKKEDKV---YIAEFKDESGEKAIKELSLKNTKLYLT 134
D 84 nkvgqgdlyvndvlyqmsaa--pasengllrtidsytsaetqetlnvlaqasvkaavegt 142
QY 135 -----YDRIFNGSAIETTPDNLKIKOIEGISVERAOKVOPMMNHAKKEIVEAID 187
D 143 qtagesygyvnyvnsfctvrvldiprkikqlgavlvtlakvypcldakansmanvqaws 202
QY 188 YLKSNAPFGKNGVNSINIDTGTDRHKAIRIDDDAKASRFKKEDL-KGTD--K 243
D 203 nyk-----ykgeglvsvlsgldprtkhmlsd--kdklksdvckftdvkh 251
QY 244 NYWLSDKIPAHFNY--NGKLTVEKYDDGRDYFDPHGHNIAGLIGAGNTEODINFGCID 302
D 252 gryfnskvpygfyndndlttdkvd-----eqghmvagllgagngtd--paksyv 303
QY 303 GIAPNAOFESKMYSDAG-SGFADETMFHAIEDSIKHNVDVSVSSGF-TGTGLVGEKY 360
D 304 gvapeagllamkvfsnctsanctsaclvsaiedsaklgadvlmslgsdngnltledpe 363
QY 361 WQAIRALKAGIPMVAVTGNATSASSSMDLVANNHLMKMTDGTGNVTRTAHEDAIIVAS 420
D 364 laavqanesslaavvsgnsgtsat--egvnhkdygldnemvsgpysgtsatlvaa 421
QY 421 AKNOTVEFDKYN-----GGESFNY--RNIGAFPKSKITTEDGTAKPSKLFVYI 470
D 422 aentdvltqavltldgltglqldpeltqlshndfctsgtdqkfyivkdasg-----n 473
QY 471 GKGDODLIGDLRGKIAVMDR-IYTKDLKNAFKKAMDGAIRIMVNTVNYNRDMTE 529
D 474 skgaladyta-dakgkialvkrgefsfdkqkyaga--agaagllivnt-----dgt 523
QY 530 LPAMGYEADGKTSQVFSISGDDGVKL--WNMINDPKKTEVKRNKKEDDFKDLQYYPID 587
D 524 tpmstalt--ctfptfglsvtgkivdwvtahpddslgk----- 563
QY 588 MESFNSKPNVGDKEIDFKAPDPTDKELYKEDIIVPAGTSWGPRIIDLKPPVSAAGK 647
D 564 -----ltlamlpn--qkyledkm--sdfstgypvanlsfkdptltpg 602
QY 648 NIKSTLN 654
D 603 niwstgn 609

RESULT 14
AAR10940
ID AAR10940 standard; Protein; 1974 AA.

```

```

XX AC AAR10940;
XX 15-APR-1991 (first entry)
XX DE Mutant protease (delta137-139/Ins15).
XX KM Mutant protease gene; fermentation: foodstuff; flavouring;
XX KM lactic acid bacteria.
XX OS Lactococcus lactis SK11.
XX FH Key Location/Qualifiers
XX FT Peptide 1..187
XX FT /Label- SIG_PEPTIDE
XX FT Protein 188..1974
XX FT /Label- MAT_PROTEIN
XX EP411715-A.
XX 06-FEB-1991.
XX PE 02-AUG-1990; 90EP-0202113.
XX PR 04-AUG-1989; 89NL-0002010.
XX PA (NEZU-) NED INST ZUIVELONDE.
XX PI Vos PAJ, Stezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX DR WPI: 1991-038622/06.
XX DR N-PSDB; AAO10870.
XX PT Mutant protease gene(s) and protease(s) - derived from type I and
XX PT III protease genes from lactococcal strains, used in fermentation
XX PT foodstuffs and flavourings
XX PS Disclosure: Fig 1(1-7)+5(b); 29pp; English.
XX CC The wild-type L. lactis SK11 protease gene sequence was determined
XX CC by the applicant (EP-307011).
XX CC The mutant protease having new cleavage specificities is
XX CC obtained by deleting three amino acids (nine bps) and inserting
XX CC 15 other residues.
XX CC This mutant may then be used to prepare hybrid proteases,
XX CC the fusion being between a type I and a type III protease of
XX CC L. lactis Wg2 and SK11.
XX CC The product has modified properties, e.g. thermostability,
XX CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
XX CC compared to the parent protease(s). The proteases can be used for
XX CC preparing products (butter cheese, human and animal foodstuffs)
XX CC prepared with the aid of lactic acid bacteria.
XX CC See also AAO10411-17 and AAO10870-71.
XX CC
XX SQ Sequence 1974 AA;

```

Query Match 12.5%; Score 431.5; DB 12; Length 1974;
Best Local Similarity 26.8%; Pred. NO. 5.5e-17;
Matches 166; Conservative 92; Mismatches 238; Indels 123; Gaps 26;

```

OY 78 NSNKSOGDYTFVNKNTKPKKEDKV--YIAEFKDESGEKAKEISLTKNTFVLYT 134
| : | | | | | : | : | : | : | : | : | : | : | : | : | : |
DB 84 nkvgqgdlyvdvlyvqnsa-pesengllrtidystaetqetnkvlaaqaavkaavegt 142
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 135 -----YDRIRGSAIETTPDNLDKIKOIEGSSVEROKQOPMNMHAKKEIGVEAID 187
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 143 qgtagesygyvngfstkryvvdiprlkqagkvctvlakvypclakansmanyavavs 202
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 188 YLKSINAPGKNPDGRCGYISNIDGTDRHKAMRIDDAKSMRPKEDL-KGTD---K 243
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 203 nyk-----ykgegtlvsvsldsgldprthkdmrlsdd--kdvkltsdvekfctlvkh 251
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

OY 244 NWLSDKIPHAENFY-NGSKITVEKYDGRDYFDPHGHIACILAGNTEODIKKFNKCID 302
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 252 gryfnskpygfyfnsadndncltdkvd-----eqhgmhvagllgangegdd--paksyv 303
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 303 GIAPNAOIFSYKMSD-----AG-----SGFAGDETFPHAIEDSIKHNVDVYSVSSG 349
| : | | | : | : | : | : | : | : | : | : | : | : | : | : |
DB 304 gvapeeqllamvfnfsnstdsgktglagktgkcgkgsavsvaleosaikgadvlnmslg 363
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 350 F-TGTGLGCEKYWQAIRALKRAGIPMVVATGVNATSSASSMDLVANNHLKMTDTGMVTR 408
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 364 snsgnqtlcdpelaavqnaesgtaavisaagsgtsast--egynvkdygldnemygs 421
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 409 TAAHEDALIVASAKNOTVEFDKVN-----GGESEFY--RNIGAFEDKSKITTTNEDG 458
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 422 pgtsgatlvaaenldvltgavltgclgqlgpetliqshdfgsgfdgkfyivkda 481
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 459 TKAPSKLKFVYIGKQDOLLIGDLRGKIAVWDR-IYTTDLNAPFKAKDKARAIAMVN 517
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 482 sg-----nlskgaladyta-dakgkialvkrgefslddkkyaqda--agaaglllvn 530
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 518 TVNYNRDNWTELPAWGYEADGRTKSQVFSIGDDGVKL--NMNINPDKTEVKKRNKED 575
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 531 t-----dgtatpmsialt--tlfptfglssvsgklvdvltahpddslygk----- 575
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 576 FKDKLEQYYPIDMESFNSKNPNVGDEKEIDFKFAPDTOKELYKEDIIVAGSTSMGPRID 635
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 576 -----ltlamlpn---qkytedkm--sdfstsygprvn 602
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 636 LLLKPDVSAPEGKNIKSTLN 654
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 603 lsfkpdltapgsnltwstgn 621
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 15
AAP94145
ID AAP94145 standard; protein; 1560 AA.
XX
XX AAP94145;
XX AC
XX 07-JUN-1990. (first entry)
XX DE S. cremoris protease.
XX KM Lactic acid bacteria; cheese; Streptococcus cremoris SK112;
XX KM protease; pK112; chymosin; prochymosin.
XX PN NL8701378-A.
XX XL
XX 02-JAN-1989.
XX PD
XX PF 12-JUN-1987; 87NL-0001378.
XX PE
XX PR 12-JUN-1987; 87NL-0001378.
XX PX
XX PA (NEZU-) NEDERL INS ZUIVELON.
XX PI
XX PT Simons AFM, De Vos WM;
XX DR WPI: 1989-030097/04.
XX DR N-PSDB; AAN91159.
XX PT
XX PT DNA fragment having region specific for lactic acid bacteria -
XX PT is contained in plasmid in microorganism used in prodn. of
XX PT protein and food prodn. eg cheese.
XX PS
XX PS Claim 6; fig 7a; 43pp; Dutch.
XX
XX The DNA encoding prochymosin can be cloned into a plasmid (esp from
XX S. cremoris SK112) and used to produce large amounts of the protein by
XX recombinant DNA techniques. The protein synthesis is driven by the
XX regulatory region of this protease. This could overcome the shortage
XX of prochymosin due to a shortage of calf stomachs and increasing cheese
XX prodn. Prochymosin is also used in prodn. of yoghurt, butter and

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: December 7, 2001, 11:45:25 ; Search time 14.18 Seconds
(without alignments)
1053.751 Million cell updates/sec

Title: US-09-590-991-6
Perfect score: 3454
Sequence: 1 PVPVLADTSSSEDALNISDK.....FGKNIKSTLVNIGKSTGY 664

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3269	94.6	1007	4	US-08-961-083-216
2	302	8.7	1164	2	US-08-589-756-1
3	302	8.7	1164	4	US-09-206-800-1
4	296.5	8.6	1150	2	US-08-589-756-3
5	296.5	8.6	1150	4	US-09-206-800-3
6	293	8.5	1167	2	US-08-589-756-2
7	293	8.5	1167	4	US-09-206-800-2
8	273	7.9	383	4	US-09-206-800-8
9	273	7.9	383	4	US-09-206-800-11
10	268	7.8	383	4	US-09-206-800-6
11	268	7.8	383	4	US-09-206-800-9
12	264	7.6	383	4	US-09-206-800-7
13	264	7.6	383	4	US-09-206-800-10
14	236	6.8	654	4	US-08-894-818B-35
15	229.5	6.6	659	4	US-08-894-818B-1
16	214.5	6.2	522	4	US-08-894-818B-5
17	201	5.8	522	4	US-08-894-818B-3
18	196	5.7	382	2	US-08-504-265B-75
19	181.5	5.3	381	1	US-07-772-087-5
20	181.5	5.3	381	1	US-08-322-965-2
21	180	5.2	382	1	US-08-460-343B-2
22	180	5.2	382	1	US-08-398-028B-2
23	180	5.2	382	1	US-08-504-265B-2
24	180	5.2	903	2	US-08-750-532-1
25	180	5.2	1398	1	US-08-750-532-9
26	180	5.2	1398	1	US-08-894-818B-8
27	177	5.1	352	1	US-07-923-260A-1

28	177	5.1	352	1	US-07-923-260A-6	Sequence 6, App11
29	177	5.1	379	1	US-08-685-774-2	Sequence 2, App11
30	175.5	5.1	378	1	US-07-772-087-4	Sequence 4, App11
31	175	5.1	379	2	US-08-845-295A-1	Sequence 1, App11
32	175	5.1	379	3	US-09-140-933-1	Sequence 1, App11
33	175	5.1	379	4	US-09-146-661-1	Sequence 1, App11
34	175	5.1	379	4	US-09-150-515-1	Sequence 1, App11
35	174	5.0	352	1	US-07-923-260A-3	Sequence 3, App11
36	173.5	5.0	350	1	US-07-923-260A-4	Sequence 4, App11
37	173	5.0	382	4	US-09-255-502-2	Sequence 2, App11
38	173	5.0	382	6	5472855-2	Patent No. 5472855
39	171	5.0	352	1	US-07-923-260A-2	Sequence 2, App11
40	171	5.0	380	6	RE34606-2	Patent No. RE34,60
41	169.5	4.9	381	1	US-07-772-087-6	Sequence 6, App11
42	169	4.9	381	1	US-08-173-508-12	Sequence 12, App1
43	169	4.9	381	2	US-08-265-510-12	Sequence 12, App1
44	169	4.9	381	3	US-08-951-742-12	Sequence 12, App1
45	168.5	4.9	380	3	US-09-049-867-2	Sequence 2, App11

ALIGNMENTS

RESULT 1
US-08-961-083-216
; Sequence 216, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 216:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1007 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-216

Query Match 94.6% ; Score 3269 ; DB 4 ; Length 1007 ;
Best local Similarity 100.0% ; Pred. No. 2.8e-209 ;
Matches 62 ; Conservative 0 ; Mismatches 0 ; Indels 0 ; Gaps 0 ;
QY 38 ETSDFFKRTAVIKREKVVSKNPVINDNNTSNEBAKIKENSNSKSGDYYDSFVNKNTEN 97
|||||

```

Db      1 ETSDFEKKRTAVIKEKEVSVKNPVIDNNTSNEAKIKEENSNNKSQGDYDTSFVNKNNTEN 60
QY      98 PKKEDKVVYIAEFKDKESGEKAIKELSLKNTKVLTYTDRIFNCSAIEPTDNLKIQOI 157
Db      61 PKKEDKVVYIAEFKDKESGEKAIKELSLKNTKVLTYTDRIFNCSAIEPTDNLKIQOI 120
QY      158 EGISSVERAOKVOPMMNHARKKEIGVEEAIDYLSKINAPFGKNFDGRGVNISNIDGTDR 217
Db      121 EGISSVERAOKVOPMMNHARKKEIGVEEAIDYLSKINAPFGKNFDGRGVNISNIDGTDR 180
QY      218 HKARIDDDAKASMRKFKEDLKGTIDKNYLSDKIPHAENYNGKITYEKTIDGGRDYDP 277
Db      181 HKARIDDDAKASMRKFKEDLKGTIDKNYLSDKIPHAENYNGKITYEKTIDGGRDYDP 240
QY      278 HGMHAGIAGLNDTEODIKNFGNIDGIAIPNAOIFSYKMYNAGSGFAGDETFMFAIEDSI 337
Db      241 HGMHAGIAGLNDTEODIKNFGNIDGIAIPNAOIFSYKMYNAGSGFAGDETFMFAIEDSI 300
QY      338 KHNVDVSVSSGFTGTGLVGEKYQWAIKRALRKAGIPMVVATGNVATSSASSSMDLVANNH 397
Db      301 KHNVDVSVSSGFTGTGLVGEKYQWAIKRALRKAGIPMVVATGNVATSSASSSMDLVANNH 360
QY      398 LKMTDGNVTRTAHEDAIAVASAKNQTVEFDKNVIGESFRYRNIGAFEDKSKITTED 457
Db      361 LKMTDGNVTRTAHEDAIAVASAKNQTVEFDKNVIGESFRYRNIGAFEDKSKITTED 420
QY      458 GTRKAPSKLFEYIGGQODLIGDLRGKIAVMDRITRKDLNAFKKAMDGARAIMVYN 517
Db      421 GTRKAPSKLFEYIGGQODLIGDLRGKIAVMDRITRKDLNAFKKAMDGARAIMVYN 480
QY      518 TVNYYNRDNTLTPAMGEADEGTRKSOVFSISGDDGVKLMNINPDKKTEYKRNKEDFK 577
Db      481 TVNYYNRDNTLTPAMGEADEGTRKSOVFSISGDDGVKLMNINPDKKTEYKRNKEDFK 540
QY      578 DKLQOYVIDMESNSKPNVNGDEKEIDFKAPDTEKLYKEDITYPAGSTWGRIDL 637
Db      541 DKLQOYVIDMESNSKPNVNGDEKEIDFKAPDTEKLYKEDITYPAGSTWGRIDL 600
QY      638 LKPDVSAPGKNIKSTLVYNGKSTYGY 664
Db      601 LKPDVSAPGKNIKSTLVYNGKSTYGY 627

RESULT 2
US-08-589-756-1
: Sequence 1, Application US/08589756
: Patent No. 5846547
: GENERAL INFORMATION:
: APPLICANT: STREPTOCOCCAL C5A PEPTIDASE VACCINE
: TITLE OF INVENTION: STREPTOCOCCAL C5A PEPTIDASE VACCINE
: NUMBER OF SEQUENCES: 5
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/589,756
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1164 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-589-756-1

```

```

Query Match      8.7%; Score 302; DB 2; Length 1164;
Best Local Similarity 24.3%; Pred. No. 4,1e-12;
Matches 146; Conservative 90; Mismatches 194; Indels 172; Gaps 32;
QY      97 NPKKEDKVVYIAEFKDKESGEKAI-----KELSLKNTKVLTYTDRIFNCSAIEPTT 147

```

```

Db      24 NAOSDIKANTVTE--DTPATEQAVETPOPTTVSEEVSSKETKTPQTPD-----DAEETV 76
QY      148 PDNLDKTIQIGISSVERAOKVOPMMNHARKKEIGVEEAID--YLSKINAPFGKNFDGRGM 205
Db      77 ADDANDL-----APQAPAKTPTDS--ATSKATIRLDLNDPSQVYTLQEKAGK--GAGT 124
QY      206 VISNIDGTDRHKKAMRIDDDAKASMRKFKEDLKGTIDKNY-----WLSDKIPHAENYNN 259
Db      125 VVAVIDAGFDKNHHEAMRLTDKAKARYQ--SKEDLEAKKHEHGTIGYEWVNDKAVAYHYDYSK 183
QY      260 GGTITVEKDYDGRDYEDPHGHIAGIAGLNDTEODIKNFGNIDGIAIPNAOIFSYKMYN 315
Db      184 DGKTRAVDQ-----EHGTHVSGILSGN--APSETEKPEYRLEGAMPPEAQLMLMVEIYN 233
QY      316 -YSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSG---FTGTGLVGEKYQWAIKRALRKAG 371
Db      234 GLADYARNYA-----QAIRDAVNLGARVINNNSFQNALAYANLPDETKKRPVYAKSK--G 286
QY      372 IPMVVATGNVATSSASSSMDLVANNHLMKMTDGNVTRTAHEDAIAVA--SAKNOTVEFD 429
Db      287 VRIYVTTAGNDSSFGKTRPLA--DH---PDYGVYGTFAAADSTLTVAISYSPDNLTEETA 341
QY      430 KVNIGSESFKYRNIGAFEDKSKITTEDGTRAPSK-LKFYIIGQODLIGDLRGKIA 488
Db      342 MYKTDQODK-----EMPLSTNR---FEPNKAQDYAAYANRGMKEDDF-KDYKGRKIA 389
QY      489 VMDRI---YTNDKNAFKKAMDGARAIMVYNVTVYVNRD-----NMTLTPAMGEA 537
Db      390 LIERSDIDFTDKIANA--KKA---GAVGVLI---YDQODGFPIELPNVQMPAA----- 436
QY      538 DEGTRKSOVFSISGDDGVKLMNINPDKKTEYKRNKEDFKLEQOYVIDMESFNSKNPN 597
Db      437 -----FISRKDGLL-----KD----- 448
QY      598 VGDEKEIDFKAPDTEKLYKEDITYPAGS-----TSWGRIDLKLPDVSAPGKNIK 650
Db      449 -NSQKTIFFNATPK-----VLPTASGTKLSRFSWGLTADGNIKPDIAGQDIL 497
QY      651 ST 652
Db      498 SS 499

RESULT 3
US-09-206-800-1
: Sequence 1, Application US/09206800
: Patent No. 6270775
: GENERAL INFORMATION:
: APPLICANT: Cleary, P. P.
: TITLE OF INVENTION: Streptococcal C5a peptidase vaccine
: FILE REFERENCE: 600 349US2
: CURRENT APPLICATION NUMBER: US/09/206,800
: CURRENT FILING DATE: 1998-12-07
: EARLIER APPLICATION NUMBER: US 08/589,756
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 1164
: TYPE: PRT
: ORGANISM: Streptococcus pyogenes, strain 49
: US-09-206-800-1

```

```

Query Match      8.7%; Score 302; DB 4; Length 1164;
Best Local Similarity 24.3%; Pred. No. 4,1e-12;
Matches 146; Conservative 90; Mismatches 194; Indels 172; Gaps 32;
QY      97 NPKKEDKVVYIAEFKDKESGEKAI-----KELSLKNTKVLTYTDRIFNCSAIEPTT 147
Db      24 NAOSDIKANTVTE--DTPATEQAVETPOPTTVSEEVSSKETKTPQTPD-----DAEETV 76

```


Db 271 LPDETKAFDYAKSGVSVTSAGNDSFSGKTRPLA--DH---PDYGVGTGTPAAADST 325
Qy 416 IAVA--SAKNOTVERFDKYNIGESFKNRINIGAFDPKSKITTNEDGTAKPSK-LKFIYICK 472
Db 326 LTVASYSFDKQLETVERVKTADQDQK-----EMPVLSTNR--FEENKAYDIAYANR 374
Qy 473 GQDDDLGLDLRGKIAVMDRIYTKDLKNAFKKAMDGARAIMVNTVNYNRD-----525
Db 375 GKEDDF-KDVKGKIALIER-GDIDFKDKIAKAKAGAVGLI-----YDNQDKGFIEL 427
Qy 526 -NMTLPMAGTEADEGTSQVFSISGDGVKLMNMINPDKTEVKRNKEDFKLEQYY 584
Db 428 PNVDMPAA-----FISRKDGLLKD--NP-----450
Qy 585 PIDMSFSNKNPNVGEDEKIDFKFAPDIDKELKEDIIVPAGS-----TSMGPRIDL 637
Db 451 -----OKTTFNATPK-----VLPTASGKTLRSFSSMGLTADGN 484
Qy 638 LKPDVSAPGKNIKSTPLANYNGK 659
b 485 IKPDIAFGODILS--SVANNK 504

RESULT 6
US-08-589-756-2
; Sequence 2, Application US/08589756
; Patent No. 5846547
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: STREPTOCOCCAL C5a PEPTIDASE VACCINE
; NUMBER OF SEQUENCES: 5
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/589,756
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1167 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-589-756-2

Query Match 8.5%; Score 293; DB 2; Length 1167;
Best Local Similarity 23.0%; Pred. No. 1.6e-11;
Matches 139; Conservative 93; Mismatches 207; Indels 166; Gaps 29;

Qy 97 NPKKEDKVVYIAEFKDKSEGEKAI-----KELSLKNTKVLTYDRINGSAIETT 147
Db 24 NAQSDIKANTYTE--DTPVTEQAVETPQPTAVSEVSSKETKTPQTPD-----DAEETI 76
Qy 148 PDLNDKIKQIEGISSEVERAKQVOPMNNHARKEIGVEAID--YLSKINAPFGKNFGDGRM 205
Db 77 ADDANDLAPQAPAKTAD-----TPATSKATIRDLNPSQVKTLOEKAK--GAGT 124
Qy 206 VISINDTGTDRHKAMRIDDAKASMRKKEDLKGTDKNY-----WLSDKIPHAFNYN 259
Db 125 VVAVIDAGFDKNEHAWRLDTKTARYQ--SKEDLEKAKKEHGIYGEVNVADKVAHYHYSK 183
Qy 260 GSKITVEYDGDGRDYDPHGHMAGILAGNTEODIKNFNGIDGIAFNAQIFSYKM----315
Db 184 DSKTAVDQ-----EHGTHVSGILSGN-APSETKEPYRLGAMPEAQLLMRVEIYN 233
Qy 316 -YSDAGSGFAGDETFMHAIEDSIKHNVNVSVSSGFTGTGLVG--EKYQAIKRALRKAGI 372
Db 234 GLADYARNYA-----QAIRDAVNLGAKVYINMFGNALALAYANLPDETKRAFDYAKSKGV 287
Qy 373 PMVVAATGNATSSASSSMDLVANNHLKMTDGNVTRTAHEDAIAVA--SAKNOTVERFDK 430

Db 288 SIYTSAGNDSFSGKTRPLA--DH---PDYGVGTGTPAAADSTLTVASYSFDKQLETETAM 342
Qy 431 VNIGSEFKNRINIGAFDPKSKITTNEDGTAKPSK-LKFVYIGQDDDLGLDLRGKIAV 489
Db 343 VKTDDQDQK-----EMPVLSTNR--FEENKAYDIAYANRKKEDDF-KDVKGKIAL 390
Qy 490 MDRITTKDLKNAFKKAMDGARAIMVNTVNYNRD-----NMTLPMAGTEADEGT 541
Db 391 IER-GDIDFKDKYANNAKAGAVGLI-----YDNQDKGFIELPNVQMPAA-----436
Qy 542 KSGVFSISGDGVKLMNMINPDKTEVKRNKEDFKLEQYYPIDMSFSNKNPNVGE 601
Db 437 -----FISRKDGLLKD--NP-----Q 451
Qy 602 KETDFKAPDIDKELKEDIIVPAGS-----TSMGPRIDLKPDVSAPGKNIKSTLN 654
Db 452 KITTFNATPK-----VLPTASGKTLRSFSSMGLTADGNIKPDIAFGODILS--S 499
Qy 655 VINGK 659
Db 500 VANNK 504

RESULT 7
US-09-206-800-2
; Sequence 2, Application US/09206800
; Patent No. 6270775
; GENERAL INFORMATION:
; APPLICANT: Cleary, P. P.
; TITLE OF INVENTION: Streptococcal C5a peptidase vaccine
; FILE REFERENCE: 600.349US2
; CURRENT APPLICATION NUMBER: US/09/206,800
; EARLIER APPLICATION NUMBER: US 08/589,756
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1167
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes, strain 12
; US-09-206-800-2

Query Match 8.5%; Score 293; DB 4; Length 1167;
Best Local Similarity 23.0%; Pred. No. 1.6e-11;
Matches 139; Conservative 93; Mismatches 207; Indels 166; Gaps 29;

Qy 97 NPKKEDKVVYIAEFKDKSEGEKAI-----KELSLKNTKVLTYDRINGSAIETT 147
Db 24 NAQSDIKANTYTE--DTPVTEQAVETPQPTAVSEVSSKETKTPQTPD-----DAEETI 76
Qy 148 PDLNDKIKQIEGISSEVERAKQVOPMNNHARKEIGVEAID--YLSKINAPFGKNFGDGRM 205
Db 77 ADDANDLAPQAPAKTAD-----TPATSKATIRDLNPSQVKTLOEKAK--GAGT 124
Qy 206 VISINDTGTDRHKAMRIDDAKASMRKKEDLKGTDKNY-----WLSDKIPHAFNYN 259
Db 125 VVAVIDAGFDKNEHAWRLDTKTARYQ--SKEDLEKAKKEHGIYGEVNVADKVAHYHYSK 183
Qy 260 GSKITVEYDGDGRDYDPHGHMAGILAGNTEODIKNFNGIDGIAFNAQIFSYKM----315
Db 184 DSKTAVDQ-----EHGTHVSGILSGN-APSETKEPYRLGAMPEAQLLMRVEIYN 233
Qy 316 -YSDAGSGFAGDETFMHAIEDSIKHNVNVSVSSGFTGTGLVG--EKYQAIKRALRKAGI 372
Db 234 GLADYARNYA-----QAIRDAVNLGAKVYINMFGNALALAYANLPDETKRAFDYAKSKGV 287
Qy 373 PMVVAATGNATSSASSSMDLVANNHLKMTDGNVTRTAHEDAIAVA--SAKNOTVERFDK 430
Db 288 SIYTSAGNDSFSGKTRPLA--DH---PDYGVGTGTPAAADSTLTVASYSFDKQLETETAM 342


```

OY 431 VNIGCESKRYKNICAFEPKSKITITNEDGTAPSK-LKFWYIGKODDGLDGRKIAV 489
Db 343 VKTDDQDCK-----EMPLSTNR--FEPNAYDIAYANRKMKEDDF-KDVGKIAL 390
OY 490 MDRITYTKDLKNAFKKAMDGCARAIMVNTVYVNRD-----NMTELPMAGYEADGT 541
Db 391 IER-GDIDFKKVKANNAKAGAVGLI-----YDQDGFPELPLPNVQMPAA----- 438
OY 542 KSOVFSISGDDGVKLMNNINPDKTEVKRNKEDFKDLEQYPIDMESFNSNKNPVGDE 601
Db 437 -----FISRKGDLLKLD--NP-----Q 451
OY 602 KEIOKFRPDTDKELYKEDITVPAQS-----TSWGPRIDLLKPYVSAFGKINXTLN 654
Db 452 KTFIFNATPK-----VLPTASGKTLRFSSWGLTADGNIKPIDIAAPGODILS--S 499
OY 655 VINGK 659
Db 500 VANNK 504

RESULT 8
US-09-206-800-8
: Sequence 8, Application US/09206800
: Patent No. 6270775
: GENERAL INFORMATION:
: - APPLICANT: Cleary, P. P.
: - TITLE OF INVENTION: Streptococcal C5a peptidase vaccine
: - FILE REFERENCE: 600.349052
: - CURRENT APPLICATION NUMBER: US/09/206,800
: - CURRENT FILING DATE: 1998-12-07
: - EARLIER APPLICATION NUMBER: US 08/589,756
: - EARLIER FILING DATE: 1996-01-22
: - NUMBER OF SEQ ID NOS: 11
: - SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 8
: LENGTH: 383
: TYPE: PR1
: ORGANISM: Streptococcus agalactiae
US-09-206-800-8

```

	Best Match	7.9%	Score 273;	DB 4;	Length 383;	
	Bust Local Similarity	24.0%;	Pred. No. 7.9e-11;			
	Matches	115;	Conservative	73;	Mismatches 136;	IndeIs 136; Gaps 23.
Oy	211	DGTDTYRHKAMRIDDDAASMRKKEDLKGTDKNY-----WLSDKIPAFNYVNGKIT	264			
		1 DAGFGKNHDAWLTGDKTARYO-SKEDELEKAKHEIGITYGEVYNNDKVAYHDYSKDGA	59			
Oy	265	VKEYIDGDGRDYDPHGCMHTAGILAGDTEODIKNFNGICDIAPNAQIFESKM-----YSDA	319			
Db	60	VDO-----EHCETHVSGLISGN-ABSETEPYRLBECAMEEAQOLLMLRVEIVNGLADY	109			
Oy	320	GSGFAGDETFMHAIDSJIKHNVDYSSVSSCGFTGTGLVG--EKYUOAIRALBRAGPIMVA	377			
Db	110	ARNYA-----QAIDATILCAKYVINMSGNAALAYANLPDETCKRAFDYAKSKGVSYIWS	163			
Oy	378	TGNVATSSASSSMDLVANNHLMKTDTGNVTRTAHEDAIAVA--SAKNQTVFEFKVNIGG	435			
Db	164	AGNDSSFEGGKTRLPLA--DH---PRYGUVGTTPAADSTLTIVASYSPDKQLTETVARVKTD	218			
Oy	436	ESFKFRNIGAFEDRSKITITNEGDTAPSR-LKFVYTGKQODDLGLDGRTIANVDRIT	494			
Db	219	QQDK-----EMEVLSTNR--PEPNAYDAYAIANKRKTEDDF-KDVKGSIALLER-G	265			
Oy	495	TKDLNAEKKAMDKGARAIMVNTVNYNYNR-----NMTELPMAGYEADEGTSQYVF	546			
Db	266	DIDFKDRIAKAKKAGAVGLI----YDNQDKGFPIELPNVDQMPPAA-----	307			
Oy	547	SLSGGDGKFLMMIMIPDKTEYKRNNKEDFKDLQGYFIIDMESFSNSKPNVNGDEKEIDF	606			
Db	308	FISRNDQGLLLK--NP-----QTTIFF	327			

Oy 607 KEADTDEKEDIIVPAG-----TSWPRIDLLPKDVSAPKNKSTLVYNGK 655
| | | | |
Db 328 NATK-----VLPTASGTRLSRFSSMGLTADGNINFDIAARQDILS--SVANNK 375

RESULT 9
US-09-206-800-11
Sequence 11, Application US/09206800
Patent No. 6270775
GENERAL INFORMATION:
APPLICANT: Cleary, P. P.
TITLE OF INVENTION: Streptococcal C5a peptidase vaccine
FILE REFERENCE: 600.349US2
CURRENT APPLICATION NUMBER: US/09/206,800
CURRENT FILING DATE: 1998-12-07
EARLIER APPLICATION NUMBER: US 08/589,756
EARLIER FILING DATE: 1996-01-22
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 11
LENGTH: 921
TYPE: PRF
ORGANISM: Streptococcus agalactiae
US-09-206-800-11

Query Match	7.98;	Score 273;	DB 4;	Length 921;
Best Local Similarity	24.08;	Pred. No. 2.5e-10;		
Matches 115;	Conservative 73;	Mismatches 156;	Indels 136;	Gaps 23

```

OY      211 DTGNDYRKAIRIDDAKASRFKKEDLKGTDKVY-----WLSDKJPRHFNVNYNGKIT 264
Db      1 DAGDKNKHEARLRDLTKRKARQ - SKEDDEKAKKKEHGITGEMVNDKAYAYHYDSKODKTA 59
OY      265 VEKYDDGHDYDPHGMHITAGLACNDTEODIKNFNGIDGLAPNAQIFSYKM-----YSDA 319
Db      60 VDO-----EHGTHVSGILSN - APSEFKPEYRLEGAMPEAQOLLMRVEIYNGLADY 109
OY      320 GSGFGDETFMHALIEDSKHNADVAVSSGPTGVLG - EKYQAOLRALRKAGIPMVA 377
Db      110 ARNYA-----QALRDALNLCAGKVINNSFGAALATAYANLPDETRKKAADYAKSKCVSIYTS 16
OY      378 TGNVATSSSSSMDLVANNHLKMTDGTGNVFTTAHEDAIAYA - SAKNQYVEFDKIGG 435
Db      164 AGNDSSFGCKRPLPLA - DH - - PDQYVGVRPAADSLFTYASVSPKQLETETRVKXTAD 218
OY      436 ESFKYRINAGAFDDSKITTTNEDGTRKAPSK - LKFYIIGKGQDODLIGLIDLRKGIAMVDRY 49
Db      219 QODK-----EMPVYSTNR - FEPRKAYDAYANNGTKEDF - KDVKQIKALIR - G 265
OY      495 TKDLKNAFKKAMDKGARAIMVNVTVNYNRD-----NMTLDPAMGYEADDEGTSQVF 548
Db      266 DIDEFDKTLAKKAKKAGAVGL - - - YNDQDKGPIELPNDQKPA - - - - - 307
OY      547 SISGDDGYKLLNMNINPDKKTVEYKRNKKEDFKOLLEOYIPIDMESFSNSKNPNVGDEKEIDF 606
Db      308 FISRDXDGLILD - NP-----OKTITF 327
OY      607 KEAPDTOKELKEDIYVAGS-----TSMGPRIDLKRPDVSAPGNKISTLNVNG 655
Db      328 NATPK-----VLPFASGTRLSRSSFSGMLTAGDNKIKFDIAAPODILS - SVANNK 375

RESULT 10
US-09-206-800-6
; Sequence 6, Application US/09206800
; Patent No. 6270775
; GENERAL INFORMATION:
; APPLICANT: Cleary, P. P.
; TITLE OF INVENTION: Streptococcal C5a peptidase vaccine
; FILE REFERENCE: 600.349052
; CURRENT APPLICATION NUMBER: US/09/206, 800

```

```

: CURRENT FILING DATE: 1998-12-07
: EARLIER APPLICATION NUMBER: US 08/589,756
: EARLIER FILING DATE: 1996-01-22
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO: 6
: LENGTH: 383
: TYPE: PRT
: ORGANISM: Streptococcus pyogenes, strain 49
US-09-206-800-6

```

```

Query Match          7.8%; Score 268; DB 4; Length 383;
Best Local Similarity 24.7%; Pred. No. 1,76-10;
Matches 118; Conservative 69; Mismatches 148; Indels 142; Gaps 25;

```

```

QY 211 DTGTDYRHKAMRIDDDAKASRFEKEDLKGTDKNY-----WLSKIPHAENYNGKAIT 264
DB 1 DAGFDKNEAMRLTDKAKARYO-SKEDLEKAKKEGITYGEMVNDKAVYHYDYSKDGKTA 59
QY 265 VEKYDDGDRYDPHGMHITAGLADNTEQDIKNFNGIDGAPNMOIFSYKM-----YSDA 319
DB 60 VDO-----EHTHVSGLISGN-APSETKEPYRLEGAMPEAOQLLMRVEIYNGGLADY 109
QY 320 GSGFAGDETMFHAIEDSIKHNVDVSVSSG---FTGTGLVEKYOQAIRALKRAGIPYV 376
DB 110 ARNYA-----QAIRDAVNLGAKVIMSGNALALAYANLPDETKKPFYAKSK-GVRIYV 162
QY 377 ATGNVATSSSSMDLVANNHLMKMTDGNVTRTAHEDAIAVA--SAKQTVPEFDKVNIG 434
DB 163 TAGNDSSFGCKTRPLA--DH---PDYGVGTAPAAADSTLVVASYSPNOQLTERAMVYTD 217
QY 435 GESFERYNIGAFEDSKITTTNEDGKAPSK-LKFVYICKGDDOGLDGLGKIAVMRI 493
DB 218 DQDK-----EMPVLTNR---FEPNKAVDYANRGMKEDDF-KDVKGKIALIERS 265
QY 494 ---YTKDLKNAFKRAMDKAGARIMVNTVNYNRD-----NMTLPAAGYEADDEGK 542
DB 266 DIDFDKIANA-KKA---GAVGVLT---YDNQDKGFIELPNVDQMPAA-----307
QY 543 SQVFSISGDDGVKLMNMINPDKTEVRKNNKEDFKDLQYYPIDMESFNSNKPVGCDEK 602
DB 308 ---FISRKDGLL-----KD-----NSQK 323
QY 603 EIDKFAPDTEKELYKEDIIVPAGS-----TSWGPRIIDLKRPDVSAPGKNIKST 652
DB 324 TITFNATPK-----VLPTASGTRLSRFSWGLTADGNIKRDIAPGODILSS 370

```

```

RESULT 11
US-09-206-800-9
: Sequence 9, Application US/09206800
: Patent No. 6270775
: GENERAL INFORMATION:
: APPLICANT: Cleary, P. P.
: TITLE OF INVENTION: Streptococcal C5a peptidase vaccine
: FILE REFERENCE: 600.349052
: CURRENT APPLICATION NUMBER: US/09/206,800
: CURRENT FILING DATE: 1998-12-07
: EARLIER APPLICATION NUMBER: US 08/589,756
: EARLIER FILING DATE: 1996-01-22
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO: 9
: LENGTH: 921
: TYPE: PRT
: ORGANISM: Streptococcus pyogenes, strain 49
US-09-206-800-9

```

```

Query Match          7.8%; Score 268; DB 4; Length 921;
Best Local Similarity 24.7%; Pred. No. 5,56-10;
Matches 118; Conservative 69; Mismatches 148; Indels 142; Gaps 25;

```

```

QY 211 DTGTDYRHKAMRIDDDAKASRFEKEDLKGTDKNY-----WLSKIPHAENYNGKAIT 264
DB 1 DAGFDKNEAMRLTDKAKARYO-SKEDLEKAKKEGITYGEMVNDKAVYHYDYSKDGKTA 59
QY 265 VEKYDDGDRYDPHGMHITAGLADNTEQDIKNFNGIDGAPNMOIFSYKM-----YSDA 319
DB 60 VDO-----EHTHVSGLISGN-APSETKEPYRLEGAMPEAOQLLMRVEIYNGGLADY 109
QY 320 GSGFAGDETMFHAIEDSIKHNVDVSVSSG---FTGTGLVEKYOQAIRALKRAGIPYV 376
DB 110 ARNYA-----QAIRDAVNLGAKVIMSGNALALAYANLPDETKKPFYAKSK-GVRIYV 162
QY 377 ATGNVATSSSSMDLVANNHLMKMTDGNVTRTAHEDAIAVA--SAKQTVPEFDKVNIG 434
DB 163 TAGNDSSFGCKTRPLA--DH---PDYGVGTAPAAADSTLVVASYSPNOQLTERAMVYTD 217
QY 435 GESFERYNIGAFEDSKITTTNEDGKAPSK-LKFVYICKGDDOGLDGLGKIAVMRI 493
DB 218 DQDK-----EMPVLTNR---FEPNKAVDYANRGMKEDDF-KDVKGKIALIERS 265
QY 494 ---YTKDLKNAFKRAMDKAGARIMVNTVNYNRD-----NMTLPAAGYEADDEGK 542
DB 266 DIDFDKIANA-KKA---GAVGVLT---YDNQDKGFIELPNVDQMPAA-----307
QY 543 SQVFSISGDDGVKLMNMINPDKTEVRKNNKEDFKDLQYYPIDMESFNSNKPVGCDEK 602
DB 308 ---FISRKDGLL-----KD-----NSQK 323
QY 603 EIDKFAPDTEKELYKEDIIVPAGS-----TSWGPRIIDLKRPDVSAPGKNIKST 652
DB 324 TITFNATPK-----VLPTASGTRLSRFSWGLTADGNIKRDIAPGODILSS 370

```

```

RESULT 12
US-09-206-800-7
: Sequence 7, Application US/09206800
: Patent No. 6270775
: GENERAL INFORMATION:
: APPLICANT: Cleary, P. P.
: TITLE OF INVENTION: Streptococcal C5a peptidase vaccine
: FILE REFERENCE: 600.349052
: CURRENT APPLICATION NUMBER: US/09/206,800
: CURRENT FILING DATE: 1998-12-07
: EARLIER APPLICATION NUMBER: US 08/589,756
: EARLIER FILING DATE: 1996-01-22
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO: 7
: LENGTH: 383
: TYPE: PRT
: ORGANISM: Streptococcus pyogenes, strain 12
US-09-206-800-7

```

```

Query Match          7.6%; Score 264; DB 4; Length 383;
Best Local Similarity 23.5%; Pred. No. 3,16-10;
Matches 113; Conservative 73; Mismatches 158; Indels 136; Gaps 23;

```

Db 164 AGDSSFGCKTRPLA--DH---PDYGVGTTPAADSTLTVASISPDKOLTEPAMVKTDD 218
 QY 436 ESEKYRNIGAFEDSKITTNEDGTAKPSK-LKEVYIGKGODDLGLDGLGKIAVMDRIY 494
 Db 219 QODK-----EMPVLSNR---FEPNKAIDYAVANRGKEDDF-KDVKGKIALIER-G 265
 QY 495 TYDLKAFFKAMDKGARAIVNTVTVYNRD-----NTELPAMGYEADGTSQYVF 546
 Db 266 DIDEKDVANAKKAGAVGLI-----YDNODKGFPIELPVDQMPAA----- 307
 QY 547 STSGDGVLMNININDKTEVRNKKDEKDLQYPIIDMSFNSNKNPVGDEKEIDE 606
 Db 308 FLSKDGLLKLD--NP-----QKTTTF 327
 QY 607 KFAPDIDKELYKEDIIVPAGS-----TSWGPRIIDLKLPDVSAPGKNIKSTLYNCK 659
 Db 328 NATPK-----VLPRASGKLSRFSSWGLTADGNIKPDIAPGODILS--SVANNK 375

SULT 13
 -09-206-800-10
 Sequence 10, Application US/09206800
 Patent No. 6270775
 GENERAL INFORMATION:
 APPLICANT: Cleary, P. P.
 TITLE OF INVENTION: Streptococcal C5a peptidase vaccine
 FILE REFERENCE: 600.349US2
 CURRENT APPLICATION NUMBER: US/09/206,800
 CURRENT FILING DATE: 1998-12-07
 EARLIER APPLICATION NUMBER: US 08/589,756
 EARLIER FILING DATE: 1996-01-22
 NUMBER OF SEQ ID NOS: 11
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 10
 LENGTH: 921
 TYPE: PRP
 ORGANISM: Streptococcus pyogenes, strain 12
 US-09-206-800-10

Query Match 7.6%; Score 264; DB 4; Length 921;

Best Local Similarity 23.5%; Pred. No. 1e-09;

Matches 113; Conservative 73; Mismatches 158; Indels 136; Gaps 23;

QY 211 DTGTDVRRHKAMRIODDAKASMRKEDLKGTDKNY-----WISDKRPHAFNYNGKIT 264
 Db 1 DAQFDKNHEAWRLTDKTKARIQ-SKEDLEKAKKEHGTTEGWYNDKVAIYHDSKDEKTA 59
 QY 265 VEKYDGRDYFDPHGMHAGILAGNDTEODIKNFNGIDGAPNAQJFSYKM-----YSDA 319
 Db 60 VDO-----EHGTHVSGILSGN-APSETKPEYRLEGAMPEAQILLMVEIVNGLADY 109
 QY 320 GSGFADDETFHAFIDSIRKINNVVSYSSGFTGTGLVG--EKYQAIARALKAGIPIVVA 377
 Db 110 ARNYA-----QAIRDAVNLGAVYINNSFGNAALAVANLPDETKKADYAKSGVSVTS 163
 QY 378 TGNVATSSASSSDVYANHLKMTDTGNVTRTAHEDAIVA--SAKNQTFEEDKVNIGG 435
 Db 164 AGNDSRSGCKTRPLA--DH---PDYGVGTTPAADSTLTVASISPDKOLTEPAMVKTDD 218
 QY 436 ESEKYRNIGAFEDSKITTNEDGTAKPSK-LKEVYIGKGODDLGLDGLGKIAVMDRIY 494
 Db 219 QODK-----EMPVLSNR---FEPNKAIDYAVANRGKEDDF-KDVKGKIALIER-G 265
 QY 495 TKDLKNAFFKAMDKGARAIVNTVTVYNRD-----NTELPAMGYEADGTSQYVF 546
 Db 266 DIDEKDVANAKKAGAVGLI-----YDNODKGFPIELPVDQMPAA----- 307
 QY 547 STSGDGVLMNININDKTEVRNKKDEKDLQYPIIDMSFNSNKNPVGDEKEIDE 606
 Db 308 FLSKDGLLKLD--NP-----QKTTTF 327
 QY 607 KFAPDIDKELYKEDIIVPAGS-----TSWGPRIIDLKLPDVSAPGKNIKSTLYNCK 659

Db 328 NATPK-----VLPRASGKLSRFSSWGLTADGNIKPDIAPGODILS--SVANNK 375

RESULT 14

US-08-894-818B-35

Sequence 35, Application US/08894818B

Patent No. 6261822

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru

APPLICANT: MORISHITA, Mio

APPLICANT: YAMAMOTO, Katsuhiko

APPLICANT: MITTA, Masanori

APPLICANT: ASADA, Kiyozo

APPLICANT: TSUNASAWA, Susumu

APPLICANT: KATO, Ikunoshin

TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300

CITY: Washington

STATE: D.C.

COUNTRY: United States of America

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/894,818B

FILING DATE: 20-MAY-1998

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03253

FILING DATE: 07-NOV-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 33285/1995

FILING DATE: 12-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Browdy, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TAKAKURA-1

TELEPHONE: (202) 628-5197

TELEFAX: (202) 737-3528

INFORMATION FOR SEQ ID NO: 35:

SEQUENCE CHARACTERISTICS:

LENGTH: 654 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-894-818B-35

Query Match 6.8%; Score 236; DB 4; Length 654;

Best Local Similarity 22.3%; Pred. No. 4.6e-08;

Matches 144; Conservative 91; Mismatches 252; Indels 158; Gaps 29;

QY 62 VIDNNTSNEBAKIRE-ENSNKSQGDYTDSPFNKNTN-NPKREKDVYIABFKKESGEKA 119
 Db 17 VVGSAVAAPKEKVEQYRNVKNGGLTLPGLFRKIQKLNPNREISTVIV--FENHREKEIA 74
 QY 120 IKELSSIKNTKVLTYTDRIINGSAIETTPNDLKIKQIEG-----ISSVEAOKVQPPAMNH 175
 Db 75 VRYL-ELMGAKVRYVY-HIIPAIADLKVDLLVIGSLTGKAKLSGVRFIQEDYKTVS 132
 QY 176 ARKEIGVEAEIDVLSINAPFGKNFDGRGAVISNIDTGTDRHAKAMRIDDAKASMRFKK 235
 Db 133 AELE-GIDESAAQVMATYV-WNLGYDSGITIGIIDIGIDASH----- 173

QY 236 EDLKTGDKNWLSDKIPHAFFNYNGKITVEKVDGDRDPDPHGMHLAGNDTEODI 295
 Db 174 PDLOG-----KVGWDFVNGRSY---PYDD-----HGCHTHVASIAGGASNG 216
 QY 296 KFNKGIDGAPNAOIFSYKMSDAGSG-----FAGDETFHAIEDSKHNVDVSVSSGF 350
 Db 217 K-----YKMGAPGAKIAGIKVLAGDAGSGSISYTIKGV---WAVDNKDKYGIKIVNLISGS 269
 QY 351 TGTGLVGEKYMQAIRALKKAGIPMVVATGN-----YANSA 385
 Db 270 SOSSDGCDLALSOAVNAAMDAGLVVVAAGNSGPNKYYTIGSPAASKVITVAGVDKIDYIT 329
 QY 386 SSSSMDLVANNHLK--MTDGTNVTRTAAHEDALIVASAKNOTVEDEKYNIGSESEFYRNI 443
 Db 330 SFFSGPFPADRLKPEYVAPGNWIIAAR-----ASGTSMGQPINDYYTAAGTSMATPHV 384
 QY 444 GAFEDKSKITTNEDGTAAPSKLFEVYIGKODDOLIGDLGKIAVMDRITYTKDLKNAFK 503
 Db 385 AGI---MALLLQAHPSWTPDKVKTALI---ETADIVKPD-----EIDAIYAGAGRVNAYK 433
 QY 504 KAMDKGARAIVVNTVYNYNRDNMTLPAWGEYEADEGTSQVFSISGDDGVK---LWMMI 560
 Db 434 -----AINT---DNIAKLVTGTIVANKSGQTHQFVYISGASFTATLYMDNA 476
 QY 561 NPKRTEV--KRNKEDPK-----DKLEQY-----YPIDMESFN 592
 Db 477 NSDDLXYLDXNGNOVDYATYAYGFEKVGYNPTDGTWIKIVVSYSGSANVQYDVVSDG 536
 QY 593 S-----KRPVNGDEKEIDFKFAPDTDKELYED-----IIVAGST 628
 Db 537 SLSPGSSPSQPEPTVDAKTFQSGDHYHYDRSDTFMTVNSGAT 581

RESULT 15
 US-08-894-818B-1
 : Sequence 1, Application US/08894818B
 : Patent No. 6261822
 : GENERAL INFORMATION:
 : APPLICANT: TAKAKURA, Hikaru
 : APPLICANT: MORISHITA, Mio
 : APPLICANT: YAMAMOTO, Katsuhiko
 : APPLICANT: MITTA, Masanori
 : APPLICANT: ASADA, Kiyozo
 : APPLICANT: TSUNASAWA, Susumu
 : APPLICANT: KATO, Ikunoshi
 : TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
 : NUMBER OF SEQUENCES: 42
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Browdy and Neimark
 : STREET: 419 Seventh Street N.W., Ste. 300
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: United States of America
 : ZIP: 20004
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/894,818B
 : FILING DATE: 20-MAY-1998
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: PCT/JP96/03253
 : FILING DATE: 07-NOV-1996
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 323285/1995
 : FILING DATE: 12-DEC-1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Browdy, Roger L.
 : REGISTRATION NUMBER: 25,618
 : REFERENCE/DOCKET NUMBER: TAKAKURA-1

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 628-5197
 TELEFAX: (202) 737-3528
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 659 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-894-818B-1

Query Match 6.6%; Score 229.5; DB 4; Length 659;
 Best Local Similarity 21.8%; Pred. No. 1,36-07;
 Matches 135; Conservative 88; Mismatches 229; Indels 157; Gaps 27;

QY 61 PYIDNNTSNEEAKIKENSNSQGDYDTSFVNK--NTENPKKE-DKVYIAEFKKEGSEK 118
 Db 28 PVRNNNAVQ-----KNYGLTFGLFKKYQRMNMNOEVDVIMFGSGYCDRDRAYK 77
 QY 119 AIKELSLKNTKVLTYTDRIFNGSAIETTPDNLDKI-----KQIEGISSVERAQ 167
 Db 78 VLR-----LMGAQVXSY-KIIPAVAVKIKARDLLIAGMIDTGYFGNTRVSGIFIOEDY 132
 QY 168 KVPQPMNHARKKEIGVEEAIDYLSINAFKNGFDGRCGVNISIDTGYRHRKAMRIDDA 227
 Db 133 KYQYDDATSVSQIGADYWNLS-----GYDGSQVVAIVDTGIDANH----- 174
 QY 228 KASMRKKEDIKGTDKNMYLSDKIPHAFFNYNGKITVEKYDGDGRDPDPHGMHLAGLA 287
 Db 175 -----PDLKG-----KVIQWYDAVNGRS--TPYDD-----QCHGTHVAGIVA 209
 QY 288 GNDT--EODIKFNFGIOSIAPNAOIFSYKMSDAGSG-----FAGDETFHAIEDSKHN 340
 Db 210 GTGSVNSQYT-----GVAPGAKLVGVKVLGADSGSVSTIAGVDWV---VONKDYG 259
 QY 341 VDVSYSVSGFTGTGLVGEKYMQAIRALKAGIPMVVATGNATSSSSMDLVANNHLKM 400
 Db 260 IRVINLSLSSQSSDGDTSLSQAVNNMADAGIYCVAAAGNSGPTMYTIGSPAAS---KV 316
 QY 401 TDTGNVTRTAHEDALIVASAKNOT-----VEBDKVNIGSESEKRYNIGAF 447
 Db 317 ITVGAVD---SNDNIASFSSRGPTADGRLEPEYVAPGVDTIAPRASGTSNG--TPINDY 371
 QY 448 DK-----SKITTNEDGTA-----PSKLKFEVYIGKODDOLIGDLGKIAVMDRI 493
 Db 372 TKASGTSMAATPHVSGVGLLIQAHPSWTPDKVKTALILETA-----DIAPKEIADIA 423
 QY 494 YTKDLKNAFKKAMDKGARAIVVNTVYNYNRDNMTLPAWGEYEADEGTSQVFSISGDDG 553
 Db 424 Y-----GAGRVVYKAIRY---DVAKLFTFTGSVADKSGATHTFPVSGATF 466
 QY 554 VK---LWMMINPDKKTEYKRNKKEDEKDLQOYYPIDMESNSNKPNGDEKEIDFKFAP 610
 Db 467 VTATLYMDTGSDDLXYLDXNGNEVDYATYAYGFEKVGYNPTAGTWIKIVVSYKGA 526
 QY 611 DTDKEIKEDIIIVAGSTS 629
 Db 527 N-----YQVD--VVSDDGLS 539

Search completed: December 7, 2001, 11:46:24
 Job time: 59 sec

Fri Dec 7 13:21:47 2001

us-09-590-991-6.ra1

Page 9

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2001, 11:45:25; Search time 19.37 seconds

(without alignments)
2611.249 Million cell updates/sec

Title: US-09-590-991-6

3454

Sequence: 1 YPVVLADTSSESDALNISDK.....PGKNIKSTLVINCKSTGY 664

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

PIR_68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	453	13.1	1946	2 JCG6032	lactocepin (EC 3.4
2	434.5	12.6	1962	2 JCG6334	lactocepin (EC 3.4
3	429.5	12.4	1902	1 B44858	lactocepin (EC 3.4
4	427.5	12.4	1902	2 S06997	lactocepin (EC 3.4
5	422.5	12.2	1902	2 B45764	lactocepin (EC 3.4
6	343.5	9.9	806	2 A41341	microbial serine p
7	298.5	8.6	757	2 C84120	subtilisin-type pr
8	293	8.5	1167	1 A35066	streptococcal Csa
9	285.5	8.3	799	2 G83753	subtilisin-type pr
10	256	7.4	731	1 A35800	subtilisin (EC 3.4.
11	250	7.2	801	2 T07617-	protease TMP - L
12	248	7.2	404	2 A44833	lactocepin (EC 3.4
13	242	7.0	754	2 T06577	subtilisin-like pr
14	239.5	6.9	1433	1 A36734	subtilisin-like pr
15	238.5	6.9	745	2 JCG6119	subtilisin-like pr
16	234	6.8	736	3 JCG7518	subtilisin-like pr
17	232	6.7	856	2 T10585	serine protease
18	231.5	6.7	816	2 T08978	serine protease
19	228	6.6	747	2 T06580	subtilisin-like pr
20	227	6.6	754	2 T06579	subtilisin-like pr
21	219.5	6.4	766	2 T48389	cucumisin-like pro
22	219.5	6.4	815	2 T00538	probable serine pr
23	211.5	6.1	761	2 T07169	subtilisin-like pr
24	208	6.0	783	2 T01015	probable subtilisi
25	206	6.0	779	2 C86335	hypothetical prote
26	202	5.8	745	2 T07184	subtilisin-like pr
27	201.5	5.8	894	2 F69730	cell wall-associat
28	196	5.7	706	2 T01351	subtilisin-like pr
29	194	5.6	635	1 A29358	cerevisin (EC 3.4.

30	191	5.5	401	2 I39974	serine proteinase
31	188.5	5.5	381	2 JH0778	subtilisin (EC 3.4
32	187	5.4	645	1 SUBSMP	serine proteinase
33	186	5.4	1905	2 T18267	multidrug resistan
34	185.5	5.4	381	1 SUBS	subtilisin (EC 3.4
35	185.5	5.4	381	2 J01487	subtilisin (EC 3.4
36	184	5.3	769	2 D86335	subtilisin (EC 3.4
37	181.5	5.3	381	1 SUBS1	hypothetical prote
38	181.5	5.3	832	1 C86431	subtilisin (EC 3.4
39	180.5	5.2	1332	2 S41552	hypothetical prote
40	179	5.2	2510	2 T28160	probable transcrip
41	178.5	5.2	1398	2 T28159	hypothetical prote
42	177.5	5.1	764	2 T05768	pyrolysin (EC 3.4.
43	176	5.1	739	2 T12964	subtilisin-like pr
44	176	5.1	779	2 T07170	subtilisin homolog
45	175.5	5.1	319	2 I39866	subtilisin-like pr
					microbial serine p

ALIGNMENTS

RESULT 1

JCG6032
lactocepin (EC 3.4.21.96) precursor [similarity] - Lactobacillus delbrueckii subsp. b

N:Alternate names: cell envelope-associated serine proteinase prp

C:Species: Lactobacillus delbrueckii subsp. bulgaricus

C>Date: 31-Dec-1996 #sequence delbrueckii subsp. bulgaricus

C:Accession: JCG6032

R:Gibbert, C.; Atlan, D.; Blanc, B.; Portaller, R.; Germond, J.E.; Lapierre, L.; Moll

J. Bacteriol. 178, 3059-3065, 1996

A:Title: A new cell surface proteinase: Sequencing and analysis of the prbB gene from

A:Reference number: JCG6032; MUID:96236017

A:Accession: JCG6032

A:Molecule type: DNA

A:Residues: 1-1946 <GI>

A:Cross-references: GB:U48487; MID:96013471; PID:AA01529.1; PID:91381114

A:Experimental source: NC001489

A:Note: neither the complete nucleic acid sequence nor the complete translation are s

C:Genetics:

C:Superfamily: lactocepin, subtilisin homology

C:Keywords: hydrolase; serine proteinase

F:1-34/Domain: signal sequence #status predicted <SIG>

F:35-1946/Product: cell surface proteinase #status predicted <MAT>

F:213-631/Domain: subtilisin homology #status atypical <SBT>

Query Match 13.1%; Score 453; DB 2; Length 1946;
Best Local Similarity 25.7%; Pred. No. 2.7e-15;
Matches 175; Conservative 114; Mismatches 261; Indels 130; Gaps 28;

QY	5	LADTSSESDALNISDKKVAENKEKHHSAMETSGDFREKKT----	AVIKEEVSK 59
DB	27	LADTSSESDALNISDKKVAENKEKHHSAMETSGDFREKKT----	AVIKEEVSK 59
QY	60	NPIYDNTSNEAKIKKEENKSGC-DYDTSFVNKNTENPKKEDKVVYIAEFKDESGER 118	
DB	87	QAVSKKNSSVRIV---SLNKAAPDHT-----SKPTGSAASAKKIDQAD--QYKQOGR 138	
QY	119	AIEELSLKNTKVLVYYDRIENGSALETTPDNLDKIKQIGISSVERAOKVOPMNHARK 178	
DB	139	VIKQVEEITGNKVRFQGYLVNAFISIDMDDDIDKVDLPQVKNVTPVKVYHP----- 191	
QY	179	EIGVEEADIDKSNAPF-GKNFDGRGNTSNDTGTGVRHKAMRIDDD-----AKASMR 232	
DB	192	---TDESADQMAQVDVQWQKLGKGBGVSTIDTGDSSHODKLDGCVSTALSKSEVE 248	
QY	233	FKKEDLKGTDKNKLWSDKIPHAFTYNGKITVEKYDDGRHVDYDPHGMHAGIAGNDTE 292	
DB	249	SDSKSL-GHGKRY--TEKYVGYNVADKNDQIV---DNCGC--EMHGOHVAGIAGAN--- 297	
QY	293	QDIKKNFGIDGICAPMAQIESYKMS--DAGSGFAGDETFPAIEDSTKHNVDVSVSSGFT 351	

298 -----GOVGVAPDAQLAMKVFSSNNAKNSGAYDDDIISAIEDSVKLGADYINNSLSGV 351
OY 352 GTGL-VGEKYMQAIRALRKAGIPVAVATGNVATSSASSSSMDLVAN---NHLKMTDQVNT 407
DB 353 SSDVSPDPQOQAAVAKASBAGYINVISAGNSGVAAGSTADGNVNTGTSSTVTPGVY 411
OY 408 RTAAHEDAIAAVSAKNGVFEFDK--NIGESFKYRN--IGAEPDKSKITTNEDGTKA-- 461
DB 412 -----PDALTVASAESKVTDTYKDELGVTFSSNSSELKA---AQTTQLESNYSVL 462
OY 463 PSKLFVYVIGKQDDDLG---LDLRGKIAVMDR--ITRKOLKNAKKAKMDKARAIMVYN 517
DB 463 TKRLKLVBMGLGAGADYAEKKAAYKGGOLAVYKRGATYFSAAVANAKA--AGAGATVYIN 520
OY 518 TVNYYNRDMMTELPMAGYEADDEGTSQVFSISGDGVKLMNINPDKTEYVRNNKEDPK 577
DB 521 -----SEDDG--LLSLSDDKTFPTLGMKADGK 547
OY 578 DKLEQYPIDEMSFNSKNPNVGE---KEIDFKFAPDPTDKELYKEDIIVPAGSTSMGPRI 634
DB 548 FKLKQOKKVRASRLKFGTALIDNSRAGKMSDF-----TSWGPPTP 586
OY 635 DLLKRPDVSAPGKNKSTLN 654
DB 587 ELDFKPEITAPGCKIYSLAN 606

RESULT 2

A32634
lactococpin (EC 3.4.21.96) precursor [similarity] - Lactococcus lactis subsp. cremoris (S
N:Alternate names: cell envelope-associated serine proteinase prtp
C:Species: Lactococcus lactis subsp. cremoris
C>Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 21-Jul-2000
C:Accession: A32634
R:Yos, P.; Simons, G.; Slezep, R.J.; de Vos, W.M.
J. Biol. Chem. 264, 13579-13585, 1989
A:Title: Primary structure and organization of the gene for a procaryotic, cell envelope
A:Reference number: A32634; MUID:89340435
A:Accession: A32634
A:Molecule type: DNA
A:Residues: 1-1962 <YOS>
A:Cross-references: GB:J04962; GB:M26310; NID:g149476; PIDN:AA03533.1; PID:g149478
A>Note: The authors translated the codon GGT for residue 1103 as Tyr
C:Keywords: part of this sequence, including the amino end of the mature protein, were deter
A:Genetics:
C:Superfamily: Lactococpin: subtilisin homology
C:Keywords: hydrolase; serine proteinase; transmembrane protein
F:1-33/Domain: signal sequence #status predicted <PRO>
F:34-187/Domain: propeptide #status predicted <PRO>
F:188-1962/Product: serine proteinase, cell-envelope-associated #status predicted <MAT>
F:208-634/Domain: subtilisin homology #status atypical <SBT>
F:1938-1955/Domain: transmembrane #status predicted <TM>

Query Match 12.6%; Score 434.5; DB 2; Length 1962;
Best Local Similarity 27.0%; Pred. No. 2.4e-14;
Matches 164; Conservative 92; Mismatches 240; Indels 111; Gaps 25;

OY 78 NSNKSQGYTSPFNKKNENKKEKVV---YIAFKKESGKRAIKLSSLKNTKLYT 134
DB 84 NKVQODDIYVIVQMSAA-PASENGILRTDYSSFAEIQOETNRYIAQAQAVKAVEQVT 142
OY 135 -----YDRIFNGSAITTPDNLDKIQEIGISSVERAKGVOPMNHAREIGVEAID 187
DB 143 QOTAGESTGYVNGFSTVRYVVDIPKQLQAGVKTVTTLAKYVTPPDAAKANSMAVQAQVWS 202
OY 188 YLKSTINAFGKNGFGRGVANISIDTGDYRHKAMRIDDAKASRKFKEDL-KGTD--K 243
DB 203 NYK-----YKGEIGVSVIDSIGDPTHKDKRLSD--KDKVLTGKSDVEKFTDVKH 251
OY 244 NYKLSDKTPHAFNY-NGKFTVEKYDDGRDYFDPHGHHIGIILAGNDETDINKNFNSID 302

DB 252 GRFYSKVPYGFENYADNNDITTDKVD-----EQGHNVAGIGANGTGDD--PAKSV 303
OY 303 GIAPPAQIEFVKMYSDAG--SGFAGDETMFHAIEDISKHNVDVSVSSGF-TGTGLYGEKY 360
DB 304 GVAPEAQLAMKVFSSNSTSAKTSATVSAIEDAKIGADVLNLSLSSNGNOTLEDP 363
OY 361 WQAIARLAKAGIPVAVATGNVATSSASSSSMDLVANNHLKMTDGNVTPTAHEDAIAVAS 420
DB 364 LAAVGNANESGAAVIAISAGNSGTSAT--ECVNNDDYGLQDENVSGPSRGAFTYAS 421
OY 421 AKNOTVEFDKNI-----GGESFKY--RNIGAFDPKSKITTNEDGTAKPSKLFVYI 470
DB 422 AENTDVIIQAVITTDGTGLQLGPEPTIQLSHDFTGSPFOKKFYIVKADSG-----NL 473
OY 471 GKQDQDGLIGDLRCKIAVMDR--ITRKOLKNAKKAKMDKGAIRAIMVYNVNRDMMTE 529
DB 474 SKGALADYTA-DAKGIATVIRGSESPDKOKYAA--AGAAGLTIYV-----DGT 523
OY 530 LPAMGEADDEGTSQVFSISGDGVKLMNINPDKTEYVRNNKEDPKLEQYPI 587
DB 524 TPMISIALT--TTPPTFGLSSVTGOKLVDWYTAHPDDSLGK----- 563
OY 588 MESEFNSKNPNVGEDEKIDFKFAPDPTDKELYKEDIIVPAGSTSMGPRI 647
DB 564 -----ITLMLPN---OKYTEDKM--SDFTSYGFVSNLSFKPDITAPG 602
OY 648 NIKSTLN 654
DB 603 NIMSTON 609

RESULT 3

B44858
lactococpin (EC 3.4.21.96) precursor [validated] - Lactobacillus paracasei (strain NCT
N:Alternate names: cell-envelope-associated proteinase prtp; serine proteinase Lp151
C:Species: Lactobacillus paracasei
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-May-2000
C:Accession: B44858; C44858; A44850
R:Holck, A.; Naes, H.
J. Gen. Microbiol. 138, 1353-1364, 1992
A:Title: Cloning, sequencing and expression of the gene encoding the cell-envelope-a
A:Reference number: A44858; MUID:92381481
A:Accession: B44858
A:Molecule type: DNA
A:Residues: 1-1902 <HOLJ>
A:Cross-references: GB:M83946; NID:g149580; PIDN:AA25248.1; PID:g149582
A>Note: sequence extracted from NCBI backbone (NCBI:112261, NCBI:112263)
C:Keywords: the source is designated as Lactobacillus paracasei subsp. paracasei
A:Accession: C44858
A:Molecule type: protein
A:Residues: 189-196 <HOL2>
R:Naes, H.; Nissen-Meyer, J.
J. Gen. Microbiol. 138, 313-318, 1992
A:Title: Purification and N-terminal amino acid sequence determination of the cell-w
A:Reference number: A44850; MUID:92226694
A:Accession: A44850
A:Status: preliminary
A:Molecule type: protein
A:Residues: 189-196 <NAE>
A:Cross-references: PIDN:AA22052.1; PID:g248666
A:Experimental source: strain NCD0 151
A>Note: sequence extracted from NCBI backbone (NCBI:94706)
C:Genetics:
C:Superfamily: Lactococpin: subtilisin homology
C:Keywords: hydrolase; serine proteinase; transmembrane protein
F:1-23/Domain: signal sequence #status predicted <SIG>
F:23-187/Domain: propeptide #status predicted <PRO>
F:188-1902/Product: serine proteinase, cell-envelope-associated #status experimental
F:208-634/Domain: subtilisin homology #status atypical <SBT>

Query Match 12.4%; Score 429.5; DB 1; Length 1902;


```

OY 356 VGEKYQOAIIRALKA---GIPWVATGNATSASSSSWDLVANNHLMKMDTGNTRTAAH 412
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 296 LNPDMATSTALDMMASEGVAVTNSGN---SGPNGM-----TVCSPGTS 337
OY 413 EDLIAVASAKNQVNEVDKVNIGESSEFKYRNIGAFPDKSKIT--TNEGCTKA--PSKIKFY 468
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 338 REAIVSGATQPLNEY-AATFGS-----YSSAKVMYGNKEDVDYALKNNKEELV 385
OY 469 YIGKGODDILGLDLRGKLIAMDR--IYTKDLNNAFKKAMDGARAIMVNTVNYNRDN 526
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 386 ENGIGAKKEGEGDLGKAVAVKRGSAIFVDKADNKKKA---GAIGWVYNNLSGEEAN 442
OY 527 WTELPMAGYEADGTSQVFSISGDDGVKLMNINPDKTEVRNNKEDFKLEQYPT 586
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 443 ---VPGMSV-----PTIKLSLEDGEKL---VSALKAGETKTTEKLTYSKAL----- 482
OY 587 DMESFMSNKNPNVDEKIDKFPDIDKELYKEDIIVPAGSTWMPRIDL-LKLPVSA 645
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 483 -----GEOVADF-----SSRGPMYMDTWIKPIDISAP 508
OY 646 GKNIKSTLVNNGKSTYGY 664
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 509 GVNIVSTIPIHPDHPHYG 527

```

```

RESULT 7
C84120
subtilisin-type proteinase (EC 3.4.21.-) BH3763 precursor [similarity] - Bacillus halodurans
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence, revision 01-Dec-2000 #text, change 19-Jan-2001
C:Accession: C84120
R: Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20263314
A:Accession: C84120
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-757 <STO>
A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BAB07482.1; GSPDB:GNOC
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3763
C:Superfamily: microbial serine proteinase vpr; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:1-20/Domain: signal sequence #status predicted <SIG>

```

```

Query Match      8.6%; Score 298.5; DB 2; Length 757;
Best local Similarity 21.6%; Pred. No. 5.8e-08;
Matches 127; Conservative 97; Mismatches 181; Indels 183; Gaps 25;

OY 96 ENK---KEDKVVYIAEFKKEGSEKAIAKLSLKNTKVLTYDRIPNGSAIETTPDNL 152
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 30 EKROLAEDDTIVIVLNRDLLETTIDVER--SIPSEGLMTR-RLNKGESFQLPREEVE 86
OY 153 KIQOEGISSVERAOKVQPMNHARKIEGVEAIDYLSINAPFGKPFDRGVAWISIDT 212
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 87 KLTIVSGVEVNDVMTYATINSAPFIAEQ---VRGLMDEGVNLTGKQKVAAYIDT 142
OY 213 GTDYRHKAMRIDDAKASMRKEDLKTGDKNYWLSDKIPHAENYNGKITVEKYDDGR 272
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 143 GIDYTH-----PDLQSSYGGYDVPD-YDD-- 166
OY 273 DYRDP-----HGMIAGITLAGNTEDODIKNFNIDIGIAPMAOLFSTKMYSDAG 320
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 167 ---DPEETIASGPRTLHGTVSGIIAAN-----GGVKVAPAELEYAVRALPGG 214
OY 321 SGFAGETMHAIEDSIKHNVDVSVSSFGTGLVGEKFWOAIIRALKA---GIPWVA 377
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 215 QGIT--EQVIAALEKAVEDVDVYNISLGTVNG---PDMFISLALDAVEGCVAVVTS 268

```

```

OY 378 TGNVATSSASSSSDVLANNHLMKMDTGNTRTAAHEDAIY-ASAKNQVPEPKVNGGE 436
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 269 NGN-----SGPNM-----TVGSPGTSKAIYSVASAPLPLNTY----- 302
OY 437 SFKYRNIGAFPDKSKITTTNEDGTRAP---SKLKFVYIGKGOODLIGDLRGKLIAMDR 492
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 303 -----LTAFGEENEISLFPSSGGLPMFAFRDLPMIDVGTGTEKEMEGVAEGKVILKR 356
OY 493 ---IYTKDLNNAFKKAMDGARAIMVNTVNYNRDNMTLPMAGYEADGTSQVFSIS 549
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 357 GWPVFEKVMH-----AVAAKARGVIIYN-----TPGPPTGM-IEGCVNIPVYSIT 402
OY 550 GDDGVKLMNINPDKTEYK---RNKKEPFKQLEQYPTIDMESRNSKNPNVDEKID 605
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 403 REGEFLEQLLELQKNKEELRTIYRKEEDF-----VALFSSRGP----- 442
OY 606 FKRAPDTEKELYKEDIIVPAGSTWMPRIDLLKLPVSAKPKNIKSTL 653
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 443 -----VHTHTWD-----VKPDVAVAGVSIIDSTII 464

```

```

RESULT 8
A35066
streptococcal C5a peptidase (EC 3.4.21.-) precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
C:Date: 13-Nov-1998 #sequence, revision 13-Nov-1998 #text, change 18-Jun-1999
C:Accession: A35066; S52539
R: Chen, C.C.; Cleary, P.P.
J. Biol. Chem. 265, 3161-3167, 1990
A:Title: Complete nucleotide sequence of the streptococcal C5a peptidase gene of Stre
A:Reference number: A35066; MUID:50153964
A:Accession: A35066
A:Molecule type: DNA
A:Residues: 1-1167 <CHB>
A:Cross-references: GB:J05229; NID:9153776; PIDN:AAA26960.1; PID:9552009
A:Note: the amino end of the mature protein was determined by protein sequencing
R: Katerov, V.; Schalen, C.; Totolian, A.A.
Mol. Gen. Genet. 245, 78-85, 1994
A:Title: Sequencing of genes within the vir regulon of Streptococcus pyogenes type M1
A:Reference number: S52535; MUID:95147851
A:Accession: S52539
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-75, 'V', '77-78, 'K', '80-83, 'PS' <KAT>
A:Cross-references: GB:S75411; NID:9914107; PIDN:AAB33264.1; PID:9914112
A:Note: In Genbank entry S75411, release 106.0, the initiation codon TTG for residue
C:Genetics:
A:Gene: scpA
A:Start codon: TTG
A:Description: specifically cleaves a Lys-Asp peptide bond near the carboxyl end of h
C:Keywords: cell wall; hydrolase; serine proteinase; tandem repeat; transmembrane pro
F:1-31/Domain: signal sequence #status predicted <SIG>
F:121-167/Product: streptococcal C5a peptidase #status predicted <KAT>
F:121-526/Domain: subtilisin homology #status atypical <SBT>
F:1029-1139/Domain: cell wall spanning #status predicted <CMS>
F:1034-1101/Region: glycine/proline-rich 17-residue repeats
F:1040-1157/Domain: transmembrane #status predicted <TM>
F:130,193,512/Active site: Asp, His, Ser #status predicted

```

```

Query Match      8.5%; Score 293; DB 1; Length 1167;
Best local Similarity 23.0%; Pred. No. 2.1e-07;
Matches 139; Conservative 93; Mismatches 207; Indels 166; Gaps 29;

OY 97 NPKKEDKVVYIAEFKKEGSEKAI-----KELSLKNTVLYTYDRIPNGSAIETT 147
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 24 NAOSDIKANTVTE--DTPVTEQAVETPQPIYAVSEEVSSKETPTQPPD-----DAEETI 76
OY 148 PDNIDKIKQLEGISSVRAKQVQPMNHARKIEGVEAID--YLSINAPFGKPFDRGM 205
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 77 ADDANDLAPQAPAKTAD-----TPATSKATIRLNDPSPQVTKLQEKAGK---GAGT 124

```

[illegible]

RESULT 9
GB3753
subtilisin-type proteinase (EC 3.4.21.-) vpr precursor [similarity] - *Bacillus halodurans*
C:Species: *Bacillus halodurans*
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 19-Jan-2001
C:Accession: GB3753
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirota
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
A:Reference number: A83650; MUID:20265314
A:Accession: GB3753
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-799 <STO>
A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA04550.1; GSPD:GB00000000
A:Experimental source: strain C-125
C:Genetics:
A:Gene: vpr
C:Superfamily: microbial serine proteinase vpr: subtilisin homology
C:Keywords: hydrolase; serine proteinase
;1-29/domain: signal sequence #status predicted <SIG>

	Query Match	8.3%	Score 285.5	DB 2	Length 799
	Best Local Similarity	21.4%	Pred. No. 2.9e-07		
	Matches 140	Conservative 96	Mismatches 196	Indels 221	Gaps 30
Oy	63	IDNNNSNEAKIKE-ENSN----	KSGSDYDTSFVAKKNENPKKEDKVYIAEFKKE	ESGE	117
Db	43	IDTSSSALETYVEIEDSIIEAKHQ-----	-OKQSNELKQARQSYIEQID-----		90
Oy	118	KAKRLSLKKNKVVYVTDRIPIGSAIETPTDNLKIKIQEIGISSVERAKQVQPMNNHAR			177
Db	91	-----LVPSSVTHHEYDFELFSGFALEPAHQISILGIDSVHVVH--VPNIYEVEYDFD			141

```

OY 178 KEIGVE-----EADIDYKASINPE-GKN-----PGRGAMVISINDIDGYRHKAMRID 224
Db 142 DEVYIEKDAYSPEMLD-----SAPFIGANDAMEAGITGESEITYAILDITGYDYTH----- 190
OY 225 DDAKASMRFEKEDLKGTDKNMYLSDKIPIHAFNYYNGCKITVEKYDGRDY--EDP----- 277
Db 191 -----PDLVHAFGCKYKGDGF-IDNNDDQETPPGCPRIET 225
OY 278 -HGMIHAGIILAGNDTEQDILKNENGIDGIANNAOIFSKYKMYSDAGSGFAGGETMHATEDS 336
Db 226 TIGTHHAGIYVAANGL-----IKGVAPDANLAIYVLGPGRGSTAG--VYAGIERA 274
OY 337 IKHNVDVVSQSGETGTGLVGEKRYWQAIRAKRAGIPVYATGNYATSASSSSWDLVANN 396
Db 275 VODGADIMNLSIGNT-----LNDDPFATSI-ALDM-AMAEK 308
OY 397 HUKMDTGVNTRIAHEDAIYAVASAKNQY-----ERDKWNIGGSESKRYNIGAFPPKSK 451
Db 309 VYAVYISNGN-----SGPNMNYIGSPQTSRDALISVGTATRLPYNKY-----KAS 350
OY 452 IYTNEDGTAPS-----KLKFVYIGKQODDILGLDJRGKIAVMDRI 493
Db 351 VETS-DGIDYPSADIMGPFSDDELLELDEGTETYEAFAGLCKPGDFEEVDYEGKIALLVRG 409
OY 494 YTKDLKNAEKKAMDKCARALIMYNYNYNYNRDMWTELPAMGYEAD-EGTSSQVFSISGDD 552
Db 410 EIPIFEKA-----ENAKAAGAVGAIYYN-----VAGVOPTYVGLAIPITIMLANED 455
OY 553 GYKLMIMNIPDKKTEYVRNNKEDFEKDLQEQYPIIDMESFNSNKNPNVDEDEKIEDFKAPADT 612
Db 456 GLKMRNELNGQNT-----VTFSEIEF 476
OY 613 DKELIKEDIIYVAGSTSMGPRIDL-LLKRPVSAPGKNIKSTLAINKSTYYG 664
Db 477 DK-LVGEYV---ADSFSSRGVPHMTIMIKIPVSAAGVAIVSTIPTHQDDPEPYG 525

```

```

RESULT 10
A:55800
cucumblin (EC 3.4.21.25) precursor - muskmelon
C:Species: Cucumis melo (muskmelon)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A55800
R:Yanagata, H.; Masuzawa, T.; Nagaoke, Y.; Ohnishi, T.; Iwasaki, T.
J. Biol. Chem. 269, 32725-32731, 1994
A>Title: Cucumblin, a serine protease from melon fruits, shares structural homology
A:Reference number: A55800; MUID:95105149
A:Accession: A55800
A:Molecule type: mRNA
A:Residues: 'LIAKSPPT', 1-731 <YAM>
A:Cross-references: EMBL:D32206
A>Note: It is uncertain whether Met-1 is the initiator or whether translation is initiated by Met-1.
C:Superfamily: subtilisin-like proteinase ag1; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F.1-10/Domain: signal sequence and amino-terminal propeptide (fragment) #status predicted
F.111-615/Product: cucumblin #status predicted <MAY>
F.111-538/Domain: subtilisin homology #status atypical <SBT>
F.616-731/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F.616-616/Cleavage site: Thr-Gly (autolytic) #status predicted

```

	Query Match	7.4%	Score 256;	DB 1;	Length 731;	
	Best Local Similarity	24.1%;	Pred. No. 8.4e-06;			
	Matches 148;	Conservative 94;	Mismatches 211;	Indels 162;	Gaps 31.	
QY	94	NTEANKKEDKVYVIA-EERDKESG---	EKAIELEL----	SLSKTKTKVLVTYTDRIENGSAIET	146	
	:	:::	: :	: :	:	:
	:	:	:	:	:	:
Db	25	DSDDGKNKIYYIKMRKLLEDPSALHNRHAMLEOVVGSTFAPESYLHTKKRSENGCAVKL		1111111111111111	84	
QY	147	TPTNDIKTIKOTIGCISVERAOKVQRIMMNAAREIGIVEAEADILKSLINAFEGANPFGRCMV			206	
	:	:	:	:	:	:
	:	:	:	:	:	:
Db	85	TEEEAEKTLASMGVVSV-----FLNE-INMLHTTTRSMDEL----	GEPITLVPRRSOV		130	

```

Oy 207 ISNIDTG---PDYHNKAMRIDDAKKAAMRKCKEDLNGTCKNYLSKRIHAENFYNGKCI 263
Db 131 ESNIVYGLDGTIMPEESPSEFDEESPPPKMGTCETSNNEFCNKKIIGARSYHIGRPI 190
Oy 264 TVERYDDGRDTEPDHGHMIGILAGNDEODIKNFNGID----GIAPNAQIFSYKMYSD 318
Db 191 SPGDVNGPRD-TNGHGHTASTAAGLVSG--ANLVGLGLGTARGVGPLARIATYAVCMN 247
Oy 319 AGSGFADGETMFAHIEDSIKHNDVYVS--PGTGLVEKTYQOALIRLRKAG 371
Db 248 DG---CSTDTILAAVDALIDGVDIISLVSGGANPRHYFDATAIS--FHAV---ERG 298
Oy 372 IPMYVATG---NATSSASSMDLVANNHLKMTDIGNVRYTAAHEDATAVASAKNOTYE 427
Db 299 ILTNSMAGNGCPNFTTASLSPW-----LLSYASTMDRKF 334
Oy 428 FDKYNIK-GESFKRNTGAF-----FDKSKTTTNEGDKAPSKLKEY 469
Db 335 VTQVQIGSGSFQGVSNITPDNOYPLVSGRDIPTMGFDKSTSRFCTDKSVNPNLLK--- 391
Oy 470 IKGQODDILDLGRKIAVMDRYRTKDLNAEKKAMDKARAIMVNVNYNRRNMT 529
Db 392 -----GRIVCEASFQ---PHEFFKSLDGAAGVLMNSNRDI--ADSY-P 430
Oy 530 LPAMGYEAD--GTSQVFSISGDDGYKLLNMNINPDKTEYVRKNNKEDFKDKLEQYYPID 587
Db 431 LPSSVLPPNDLALRLRIYSIR--SPGATIF-----KSTYI-----LMSAPV- 471
Oy 588 MESFNKPNVNGDEKEIDFPAPDIDELKEDIIVPAGSTSMGPRIDLK-PDVSAG 646
Db 472 VVSFSSKCPN-----RAYDAVLIKPDIS-----GpGVETLAMPSPVAPYG 510
Oy 647 KNIKSTL-NVINGKS 660
Db 511 GIRNTLENIISGTS 525

```

RESULT 11
T07617
proteinase TMP - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999
C:Accession: T07617
R:Rt99s. C.D.
submitted to the EMBL Data Library, November 1996
A:Description: Antisense inhibition of a tomato meiotic proteinase.
A:Reference number: Z16054
Accession: T07617
Status: Preliminary; translated from GB/EMBL/DBJ
Molecule type: DNA
A:Residues: 1-801 <RIG>
A:Cross-references: EMBL:U080583; NID:g1732366; PIDD:AA838743.1; PID:g1732366
A:Experimental source: strain UC82b
C:Genetics:
A:Introns: 41/2; 91/3; 229/2; 261/2; 284/3; 379/1; 490/3; 572/1
A:Superfamily: subtilisin-like proteinase ag12; subtilisin homology

	Query Match	7.2%	Score 250;	DB 2;	Length 801;
	Best Local Similarity	22.5%;	Pred. No. 1.9e-05;		
	Matches 141:	Conservative	90;	Mismatches 237;	Indels 158; Gaps 31.
Oy	88	DSEVKNTEPNKPEKKQVVYIAEFKKESEKAIKELSSLNKKNV---LYTVDRINGSAI	144		
	: :	:	:	:	:
Db	33	DPFV--STESKLELVNIIYKERMRQHD-----MLGSLLEKSVYTKKYSYHLINFGFI	85		
Oy	145	ETTPDN-LDKIQIGIEISSVERAOKVOPMMNHARKIEIGEVAIDILKLSINAEPFKNFDR	203		
	: :	:	:	:	:
Db	86	HLTSEALDVLRLNVEGVRAIYEDVAKMKLTTHTPDLGL--PVGVMPKICGP---TTSQA	140		
Oy	204	GMAVISNIDTGDIYRIKKAM--RIDDAAKASMRKKEDLKCTKDKNYLSDKIIP-----	253		
	: :	:	:	:	:
Db	141	GVAGIMIDGTGINPFPSPLAQASNAGRGITVYSGKFPG--KCIVIGREFPTACNSKIV	197		

[illegible]

RESULT 12

A:Accession: A44833

lactocepin (EC 3.4.21.96) [similarity] - Lactococcus lactis subsp. lactis (strain UC310)

N:Alternate names: cell envelope-associated serine proteinase prtp

C:Species: Lactococcus lactis

C:Date: 03-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 19-May-2000

C:Accession: A44833; B44833

R:Law, J.; Vos, P.; Hayes, F.; Daly, C.; de Vos, W.M.; Fitzgerald, G.

J:Gen. Microbiol. 138, 709-718, 1992

A:Title: Cloning and partial sequencing of the proteinase gene complex from Lactococcus

A:Reference number: A44833; MUID:92268851

A:Contents: subsp. lactis UC317, plasmid pCJ310

A:Accession: A44833

A:Molecule type: DNA

A:Residues: 1-404 <LAW>

A:Cross-references: GB:M85288

A>Note: The authors translated the codon TAA for residues 83 and 170 as Glu and CAG f

A>Note: The sequence shown follows the authors' translation at positions 83 and 170

A>Note: sequence modified after extraction from NCBI backbone

A>Note: sequence extracted from NCBI backbone (NCBIN:103657, NCBIN:103659, NCBI:P:1036

C:Genetics:

A:Genome: plasmid

C:Superfamily: lactocepin; subtilisin homology

C:Keywords: hydrolase; serine proteinase

	Query Match	7.28;	Score 248;	DB 2;	Length 404;	
	Best Local Similarity	28.4%;	Pred. No. 9.2e-06;			
	Matches	95;	Conservative	56;	Mismatches 134;	Indels 50; Gaps 15.
OY	246	WLSDDIPHAFFNY-NGGRITVEKYDGRPRPHOMHLAGLNGNDEEDIKNFNCIDSI	304			
	:	: : : : :	:	:	:	:
Dd	1	YFNSVYPGFENADNDNTTDDKVD----	BQHMHVAGIIGANGTGDD--PANSVGV	52		
OY	305	APNAQIFSKYMYSDAG-SQFAGEETMFHAIEDSIKHNVDSVSSGF-PGTGLVGKEYWQ	362			
	:	: : : : :	:	:	:	:
Dd	53	AREADLLMKVFYTNDSITSATTVSALIVSAIEDSAKITGADVILMWSJCSDSGNOTLEDPELA	112			
OY	363	AIRALRKAGIPMVAVTGNVATSSASSSMDLVANNHLKMTDTGNVTRTAHEDAIVASAK	422			

```

Db 113 AVQANESGTAIVASNGSGTSGAT--EGVKNKYGLQDNEMVGTSGTSGATVAAAE 170
Oy 423 NOTVEFDKVN-----GGESEFV--RNTGAFPFDSKITTNEGTAPSKLKVYIGK 472
Db 171 NTDVITQAVTTTDTGTGLQGLPETIOLSSNDFTSGDOKRFVYVADASGNLSK-----GK 224
Oy 473 GQDDLLGLDGRKIAVMDRIYTKDLKNAFKAMDKARAIMVY-----VTVYNYNRDN 526
Db 225 VADT---ADAKGKIATYKR---GELTFADKQ---KYAQAGTGPLLKNKNTGTGYIGSM 275
Oy 527 WTELPAMGYEADGKTSQVFSISGDDGVKLMNMIN 561
Db 276 VTDA-----DANOIYDDQAIASFSSDKNA-LYNEIS 304

```

RESULT 13

```

T06577
subtilisin-like proteinase (EC 3.4.21.-) - tomato
N:Alternate names: proteinase p69c
C:Species: Lycopersicon esculentum (tomato)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C:Accession: T06577
R:Melchior, J.; Amrhein, N.; Schaller, A.
submitted to the EMBL Data Library, May 1998
A:Description: The gene family of subtilisin-like proteases in tomato.
A:Reference number: 215771
A:Accession: T06577
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-754 <MEI>
A:Cross-references: EMBL:AJ005171; PIDN:CAA06412.1
A:Experimental source: cultivar VFM8
C:Genetics:
A:Gene: p69c
C:Superfamily: subtilisin-like proteinase ag12; subtilisin homology
C:Keywords: hydrolase; serine proteinase

```

Query Match 7.0%; Score 242; DB 2; Length 754;

Best Local Similarity 22.58; Pred. No. 4.6e-05;

Matches 140; Conservative 92; Mismatches 236; Indels 154; Gaps 28;

```

Oy 83 QGDYDSFVNKNTENPKKEDRVYIAEFKDES-----GEAKELSLKNTKVLVT 134
Db 21 QSDDETYIV--HVESP--ESLITQSSFMDESYLSFLPTMKALISSGNEAASITYS 76
Oy 135 YDRIFNGSAITTPDNLKIKQIEGISSVERAKVQPMNHARKEIGVEAIDYLSINA 194
Db 77 YHNWMTGFAARLTAEQVEMEKKGFSVSAQKRLISLHTTTPSFLIQKNGVWKDSNY 136
Oy 195 PFGKNFDRGVNISNDGTGYRHKAMRIDDAKASMKFKEDLKTGDKNW---LSKI 251
Db 137 -----GKGVITGLVDLGIIPDHSF-----SDVGMPSPAKMGVCKSNFTKCNKL 184
Oy 252 PHAENYNGKITVEKYDDGDYDFPHGMHAGILAGNDTEODIK-----NFGCID-GI 304
Db 185 IGASVYELGNASPIDN-----DGHGTHASTAG-----AFYKGAHVHGNANGTANGV 232
Oy 305 APNAQISFYKATSDAGSFAG---DETFHAIEDSIRKHNVDVYSSGFTGTGLGEKYW 361
Db 233 APLAHIAIKYV-----CGFDKCPGSDILAMDAIDGVDILISLGSLSPLYDETIA 287
Oy 362 QAIRALKRAGIPVYVATGNTVATSSSSSSMDLVANNHLKMTGTGVTTRAHEDAIAVSA 421
Db 288 LGAVSTTGRLVSCSAGNSGSPAS-----VDNSAPILITVVGASTLDRTKATVKKLNG 342
Oy 422 KNOTVEFDKVNIGESF---KYRNIGAF--FDKSKITTNEDGTAPSKLKVYIGKGDQ 476
Db 343 E-----EFG-----GESAVHPTSNATFTLEDAK-----NANDPSTPYCRGRGSLDP 387
Oy 477 DLIGLDLGRKIAV---MDRIYTKDLKNAFKAMDKARAIMVYVTVN---YNRDNWTELP 531

```

```

Db 388 -----AIRKGIIVLCIAFGVANVDKQAVK---DAGVGMIIVNPQYGVTKSADAHLP 439
Oy 532 AMGYEADGKTSQVFSISGDDGVKLMNMINP-----DKKTEVKNKNEDEFDK 579
Db 440 ALVVSADGTKIRAVTNS-----ILNPVAITTFGCTIIGDKNAFI----- 479
Oy 580 LEQYYPIDSEFSNKNRPVNGDEKIDFPKAPDTKELYKEDIIVAGSTSGMPRIDLLK 639
Db 480 -----VAAFSSRGPT-----ASNGILKPOLI-----GPGVNIILAA 510
Oy 640 PDVSAFG-KNIKSTLVNIGKS 660
Db 511 WPTSVGKNKNTKSTFNITSGTS 532

```

RESULT 14

```

A36734
Bacillopeptidase F (EC 3.4.21.-) precursor bpr [validated] - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 16-Jun-2000
C:Accession: A36734; A35131; A35750; S08223; JN0335; I39849; B69596; JU0084
R:Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.
Bacteriol. 172, 5520-5521, 1990
A:Reference number: A36734; MUID:90368623
A:Contents: erratum
A:Accession: A36734
A:Molecule type: DNA
A:Residues: 1-1433 <SILO>
A:Cross-references: GB:M29035; NID:g143307; PIDN:AAA62679.1; PID:g143308
R:Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.
Bacteriol. 172, 1470-1477, 1990
A:Title: Bacillopeptidase F of Bacillus subtilis: purification of the protein and c
A:Reference number: A35131; MUID:90170864
A:Accession: A35131
A:Molecule type: DNA
A:Residues: 1-365, 'S', 367-682, 'EIMP', 893, 'Q', 895-896 <SL2>
A:Cross-references: GB:M29035
A:Note: the authors translated the codon GAA for residue 545 as Leu
R:Wu, X.C.; Nathoo, S.; Pang, A.S.H.; Carne, T.; Wong, S.L.
J. Biol. Chem. 265, 6845-6850, 1990
A:Title: Cloning, genetic organization, and characterization of a structural gene en
A:Reference number: A35750; MUID:90216713
A:Accession: A35750
A:Molecule type: DNA
A:Residues: 1-392, 'V', 394-828, 'NIRRLYSLKFCRSRHSV' <MUA>
A:Cross-references: GB:J05400; NID:g142607; PIDN:AAA83363.1; PID:g119197
A:Note: this sequence has been corrected
A:Accession: B35750
A:Molecule type: DNA
A:Residues: 876-935, 'CG' <MU2>
A:Cross-references: GB:J05400; NID:g142607; PIDN:AAA83363.1; PID:g119197
R:Masuda, E.S.; Anaguchi, H.; Sato, T.; Takeuchi, M.; Kobayashi, Y.
Nucleic Acids Res. 18, 657, 1990
A:Title: Nucleotide sequence of the sporulation gene spoIIIGA from Bacillus subtilis.
A:Reference number: S08223; MUID:90174995
A:Accession: S08223
A:Molecule type: DNA
A:Residues: 1410-1433 <MAS>
A:Cross-references: EMBL:X17344; NID:g40165; PIDN:CAA35224.1; PID:g809661
R:Kato, T.; Yamagata, Y.; Arai, T.; Ichishima, E.
BioSci. Biotechnol. Biochem. 56, 1166-1168, 1992
A:Title: Purification of a new extracellular 90-kDa serine proteinase with isoelectri
A:Reference number: JN0335; MUID:93005071
A:Accession: JN0335
A:Molecule type: protein
A:Residues: 195-218, 'A' <KAT>
A:Note: source of this material was Bacillus subtilis (natto)
R:Beall, B.; Lowe, M.; Lutkenhaus, J.
J. Bacteriol. 170, 4855-4864, 1988
A:Title: Cloning and characterization of Bacillus subtilis homologs of Escherichia c
A:Reference number: I39846; MUID:89006108
A:Accession: I39849

```

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-211 <RES>
A:Cross-references: GB:M22630; NID:g142938; PIDN:AAA22458.1; PID:g551705
R:Kunst, F.; Ogatawara, N.; Mosser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertea
C.: Brion, S.; Broillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Ch
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Mature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funo, S.; Gallizzi, A.; Galle
lech, J.; Harwood, C.R.; Heman, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinou
A:Authors: Lauder, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y., M.; Ogawa, K.; Ogihara, A.; Oudaga, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetell
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadle, V.; Sato, T.; Scanlon
A:Authors: Schlecht, S.; Schroeder, R.; Scoffone, F.; Sekiuchi, J.; Sekowska, A.; Sero
keuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Zechlin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
Reference number: A69580; MUID:98044033
Accession: B69596
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1433 <RUN>
A:Cross-references: GB:299111; GB:299112; GB:AL009126; NID:g2633902; PIDN:CAB13404.1; P
A:Experimental source: strain 168
A:Genetics:
A:Gene: bpr; bpf
A:Map position: 135 (degrees)
C:Superfamily: bacillopeptidase F; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine proteinase
F:1-30/Domain: signal sequence \$status predicted <SIG>
F:31-194/Domain: propeptide \$status predicted <PPO>
F:195-1433/Product: bacillopeptidase F \$status experimental <MAT>
F:218-466/Domain: subtilisin homology <SBR>
F:227,274,452/Active site: Asp, His, Ser \$status predicted

```

OY      403 TGNVTRTAHEDDALVAWSAKNOTVEE--DKVNIGGESPKRYNICAFFEDRSKITNEEDGT- 459
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      467 QANNSLSVDMEDEDLITSTAEPLDTPSPDSNNNG---YGHGLVAFPDAVSAT---DQLG 520

OY      460 KAPSLKFLVYIGKGODDD-----LIGDLRCKIAVMDRITYTKDKLNARFKAMD 507
          ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      521 KAEGQSVY----EGDDGPYPYOHEKVTEAVEGSGSLPTLTAEADNVSVTSVKLSYK--LD 574

OY      508 KGAAIIMVTVNYVNYNRNMVELPRAMGYEADEBGRKSQVFYSISGDGVLYMMKNINDKTE 567
          :||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      575 QG-----EWETITPAK-----RTSGDHLLKGTGYAEIIPDIK-- 603

OY      568 VKRNKKEDFKDLKY--YPIDMESFN-SNKPYNVGEKEIDFKAEPDPDKELYEDIIVP 624
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      604 ---GTKLSYKMKMHDFGHHVSDVDYTVKPSITAGYKORPEFAAPGC-----WVA 651

OY      625 AGSTS--WGPRIDLCLKPDVSAPEKNIKSYLVAINCKSTYG 663
          :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      652 SGTNNMEMG-----VPSTGPN-----TAASCEKVVYG 678

RESULT   15
JC6119
subtilisin-like proteinase (EC 3.4.21.-) - tomato
N:Alternate names: calcium-activated anopeptidase; proteinase p69
C:Species: Lycopersicon esculentum (tomato)
C>Date: 23-Mar-1997 #sequence_revision 09-May-1997 #text_change 20-Jun-2000
A:Accession: JC6119
R:Tornero, P.; Conejero, V.; Vera, P.
Proc. Natl. Acad. Sci. U.S.A. 93, 6332-6337, 1996
A>Title: Primary structure and expression of a pathogen-induced protease (PR-P69) in
A:Reference number: JC6119; MUID:96270538
A:Accession: Jc6119
A:Molecule type: mRNA
A:Residues: 1-745 <TR>
A:Cross-references: EMBL:X95270; NID:g1524114; PIDN:CAA64566.1; PID:g1524115
C:Comment: This enzyme, a secreted calcium-activated endopeptidase, is a plant pathog-
fense against attacking pathogens.
C:Genetics:
A:Gene: p69
C:Superfamily: subtilisin-like proteinase ag12; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:137-546/Domain: subtilisin homology #status atypical <SBT>

Query Match           6.9% Score 238.5; DB 2; Length 745;
Best Local Similarity 22.5%; Pred. No. 6.8e-05;
Matches 148; Conservative 87; Mismatches 223; Indels 199; Gaps 33;

OY      61 PVIDNN-----TSNEERAKITEENSNKSGQDYDPSFVNKTMDENPKKEDKVVYIAEFKD 113
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      18 PTIGSNETTYLVHWESPESLISTOSSLTDLDSYLSLPKTT----- 59

OY      114 ESGERAIKELSLNTKVLYTDRIFGNSAIETPPNDILDKIKOLEGISSVERAQVOPPM 173
          ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      60 ----TALSSSNEEAFTMYISYHNMTGFAARLAEOVKEMEKINGFVSAQRKRLSIDT 115

OY      174 NHARKIEIVEBAIDYLKSIAPFCGKNFDGKMVINSIDTGTDYRHKKAMRIDDAKASKRF 233
          :||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      116 THTSSTFLQDNMCMVKMSNY-----GKCVIIIGVIDTGLIRPHPSF-----SPVGMP 163

OY      234 KKEDLKGF-DKNYM--LSDKIRHAFNYNGKRTYEKDDCRDRDFDPHGMIHAGLAGN- 289
          ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      164 PRAMWKVCESNFINKCNKLIGARSYOLHGSPID-DDG-----HGHTASTAACAF 215

OY      290 -DTBDLIKFNFG-IDGIAPNAQIFSYKMYSDAGSGFACDETTFMAIEDSIKHNDVSVS 347
          :||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      216 VGNANVFENANGTAGAVAPFAHIAVYKVCNSDG---CADVDVLAAMDAADIDGVDILSIS 272

OY      348 SGFPGTGLVGEKYQDAIR---ALRKACIPMVYATGNATATSASSSSMDLVANNHLKMTDT 403
          :||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      273 LGGGGS---SDPYSNPIALGAYSATERGILVSCSGANNCPSTGS-----V 314
```

```

QY 404 GN-----VTRTAHEDAIYASAK-NQTFEDKVNIGGESEFKYR-----NIGAFEDKSK 451
Db 315 GNEAPWILTVGASTQDRKLTATVKLGNNREEFE-----GES-AYRKISNSTFFALFDAGK 368
QY 452 ITTNEEDGTRKAPSKLEKVEYIGKGDDDLIGDLRGKIAY-----MDRIYTKDLKNAFKKA 505
Db 369 NASDEFFETP-----YCRSGSLTDPV--IRGRIYICLAGGVPRV--DKGOAVK-- 412
QY 506 MDKGARAIMVNTVNYVNRDWNTE-----LPAMGYEADGEGTKSGVFISGDDGVKLMNM 559
Db 413 -DAGGVGMIITIN---OQRSGVTKSADAHVLPAL-----DISDADGTKILAY 454
QY 560 INPDKTEVKRRNNKEDFEKDLQYYPIDMESFNSNKPWGDKEIDFKFAPDTELYKE 619
Db 455 MNSTSN-----PVATITFGGT--IIGD-----KNAP----- 478
QY 620 DIYPAGSTSWGPRIDL--LKPDVSAPG-----KNIKSTLVNINGS 660
Db 479 --IYAASFSSRSGASIGILKPDILIGPGVNILAMPSTVDONKNTKSTFNISGTS 532

```

Search completed: December 7, 2001, 11:48:40
 Job time: 195 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2001, 11:46:26 ; Search time 13.44 Seconds
(without alignments)
1811.417 Million cell updates/sec

Title: US-09-590-991-6

Perfect score: 3454

Sequence: 1 YPVVLADTSSSEDALNISDK.....PGKNIKSTLVINCKSTYGY 664

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	434.5	12.6	1902	1 P3P_LACLC	P15592 lactococcus
2	429.5	12.4	1902	1 P2P_LACPA	Q02470 lactobacilli
3	427.5	12.4	1902	1 P2P_LACLC	P15293 lactococcus
4	421.5	12.2	1902	1 P1P_LACLC	P16271 lactococcus
5	343.5	9.9	806	1 SUBV_BACSU	P29141 bacillus su
6	293	8.5	1167	1 SCAL_STRPY	P15526 streptococ
7	292	8.5	1181	1 SCA2_STRPY	P58099 streptococ
8	239.5	6.9	1433	1 SUBF_BACSU	P16397 bacillus su
9	201.5	5.8	894	1 WPRB_BACSU	P54423 bacillus su
10	194	5.6	635	1 PRTB_YEAST	P09332 saccharomyc
11	191	5.5	401	1 THES_BACSP	Q45670 bacillus sp
12	188.5	5.5	381	1 SUBN_BACNA	P15833 bacillus su
13	187	5.4	645	1 SUBE_BACSU	P16396 bacillus su
14	186	5.4	1905	1 TAGB_DICDI	P54683 dictyostell
15	185.5	5.4	381	1 SUBT_BACSA	P00783 bacillus su
16	185.5	5.4	381	1 SUBT_BACST	P29142 bacillus su
17	181.5	5.3	381	1 SUBT_BACST	P04189 bacillus su
18	180.5	5.2	1332	1 SPRT_YEAST	P35177 saccharomyc
19	175.5	5.1	319	1 ISPL_BACSU	P11018 bacillus su
20	175	5.1	379	1 SUBT_BACLI	P00780 bacillus su
21	173	5.0	382	1 SUBT_BACAM	P07596 lactococcus
22	170	4.9	682	1 NISP_LACLA	Q07596 lactococcus
23	168.5	4.9	380	1 ELVA_BACAO	P27693 bacillus al
24	167	4.8	420	1 SUBT_BACSU	P28842 bacillus sp
25	163.5	4.7	380	1 ELVA_BACCS	P41362 bacillus su
26	160.5	4.6	326	1 ISP_PABPO	P29139 paenibacilli
27	159	4.6	534	1 PROA_VIBAL	P16588 vibrio alga
28	159	4.6	1296	1 ASAI_ENTFA	P17953 enterococcu
29	158	4.6	378	1 ELVA_BACSP	P20724 bacillus sp
30	157.5	4.5	1304	1 COLA_CLOPE	P43153 clostridium
31	156	4.5	2869	1 RBPL_PLAVB	Q00798 plasmodium
32	155.5	4.5	275	1 SUBT_BACPU	P07518 bacillus pu
33	153	4.4	710	1 NECH_HYDAT	P29145 hydra atten

34	152	4.4	361	1 ELVA_BACHD	P41363 bacillus ha
35	152	4.4	944	1 NUF1_YEAST	P32380 saccharomyc
36	152	4.4	960	1 YMX6_YEAST	Q04279 saccharomyc
37	151.5	4.4	1167	1 CACA_HELPJ	Q921t1 helicobacte
38	151	4.4	533	1 PEPC_ASPNC	P33295 aspergillus
39	151	4.4	1658	1 YM67_YEAST	Q0361 saccharomyc
40	150.5	4.4	802	1 NAB3_YEAST	P38996 saccharomyc
41	149	4.3	1053	1 SLPN_BACBR	P06546 bacillus br
42	146.5	4.2	1186	1 CACA_HELPY	P55980 helicobacte
43	146	4.2	477	1 PRCA_ANASO	P23915 anabaena sp
44	146	4.2	578	1 LIPA_MYCPU	Q50274 mycoplasma
45	145.5	4.2	1375	1 GTFC_STRMU	P13470 streptococc

ALIGNMENTS

RESULT 1

ID	P3P_LACLC	STANDARD	PRT	1902 AA
AC	P15292			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-APR-1990 (Rel. 14, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	PII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-ASSOCIATED SERINE PROTEINASE).			
GN	PRTP.			
OS	Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).			
OG	Plasmid.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Lactococcus.			
OX	NCBI_TaxID=1359;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.			
RC	STRAIN-SK11;			
RA	MEDLINE=89340435; PubMed=2760036;			
RX	Vos P., Simons G., Slezem R.J., de Vos W.M.;			
RT	*Primary structure and organization of the gene for a procaryotic, cell envelope-located serine proteinase.*;			
RL	J. Biol. Chem. 264:13579-13585(1989)			
CC	-1- FUNCTION: PROTEINASE WHICH BREAKS DOWN MILK PROTEINS DURING THE GROWTH OF THE BACTERIA ON MILK.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE SUBTILASE FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL: J04962; AAA03533.1; ALT_SEQ.			
CC	PIR: A32634; A32634.			
DR	HSSP: P00782; 2SBR.			
DR	MEROPS: S08_019; -.			
DR	InterPro: IPR001899; Gram_pos_anchor.			
DR	InterPro: IPR003137; PA.			
DR	InterPro: IPR000209; Peptidase_S8.			
DR	Pfam: PF00746; Gram_pos_anchor; 1.			
DR	Pfam: PF02225; PA; 1.			
DR	Pfam: PF00082; Peptidase_S8; 3.			
DR	PRINTS: PR00723; SUBTILISIN.			
DR	PROSITE: PS00136; SUBTILASE_ASP; 1.			
DR	PROSITE: PS00137; SUBTILASE_HIS; 1.			
DR	PROSITE: PS00138; SUBTILASE_SER; 1.			
DR	PROSITE: PS00343; GRAM_POS_ANCHORING; 1.			
KW	Hydrolase; Serine protease; Cell wall; Zymogen; signal; plasmid; Transmembrane.			
KW	Transmembrane.			
FT	SIGNAL 1 33			
FT	PROPEP 34 187			

```

FT CHAIN 188 1902 PII-TYPE PROTEINASE.
FT DOMAIN 188 1876 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1877 1895 MEMBRANE ANCHOR (POTENTIAL).
FT DOMAIN 1896 1902 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 1902 AA; 200550 MW; 87CECBA9345FE9D3 CRC64;

Query Match 12.6%; Score 434.5; DB 1; Length 1902;
Best Local Similarity 27.0%; Pred. No. 5,96-14;
Matches 164; Conservative 92; Mismatches 240; Indels 111; Gaps 25;

Y 135 -----YDRFNGSAIETTPDNLDKIQEGISSVERAOKVOPMMHARKEIGVEAID 187
Db 84 NKVQODIYVIVYQMSAA-PASENGILRTDYSTAEIQOETNKYIAQASVKAVEQVT 142
Y 78 NSNKSQGYTSPFNKNTEENKKEKVV---YIAEFKKEGGERAKIKLSIKTKVLYT 134
Db 84 NKVQODIYVIVYQMSAA-PASENGILRTDYSTAEIQOETNKYIAQASVKAVEQVT 142
Y 143 QQTAGESTGYVNGESTYVRVVDIKQLAGVKTLYLAKYIYPTDAKANSNAVQAVWS 202
Db 188 YLKSINAFEGKNFDGRGVNISIDTGDYRHKAMRIDDAKASMRKKED--KGTD--K 243
Y 203 NYK-----YKGEYTVSVISDIDPTHKMRLSD--KDYGLTKSDVEKFTDVKH 251
Db 244 NYWLSDKTIPAFNY--NGKITVEKYDGRDYFDPHGHINGILAGNTEDDINKNFID 302
Db 252 GREYFSKVPYGFENYADNDITTDKVD-----EQHGMHAGIIGANTGDD--PAKSV 303
Y 303 GIAPNAQIFSKYKMSDAG--SGFAGDETFMFAIEDSIKHNVVSVSSGF--TGTLVGEKY 360
Db 304 GVAPEAOQLAKKAVNSNSTSAKTSATVSAIEDSAKIGADVLNLSGNSNGNLTLEPE 363
Y 361 WOAIRALKKAGIPVNAVATGNAVATSASSSSMDLVANNHLKMTDGTGNTVTAHEDAIAVAS 420
Db 364 LAAVONANESGTAIVISAGNSGTGSAT--EGVNDKYGLQDNEKVSFGTSRGATVVAS 421
Y 421 AKNOTVERDKNI-----GGSFKY--RNIGAFPKSKITINEDCTAPSKIKFYI 470
Db 422 AENDVITQAVTITDGTGLQLGPEITQLSSHDFTGSGFPOKRFYLVKASG-----NL 473
Y 471 KKGODDGLIGLNGKIAVMDR--IYTKDLKNAFKAMDKARAIIVTVYVNRDMTE 529
Db 474 SKGALADYTA--DAKGKIAIVKRGESFPDKOKYQA--AGAAGLIIVT-----DSTA 523
Y 530 LPAMGYEADGCKSOVESISGDDGVKL--WMMINPDKKTEVYKRNKEDFKDKLEQYYPID 587
Db 524 TPMTSIALT--TTEPTFGLSSVTGQKLVADVTAAHPDDSLGVK-----563
Y 588 MESFNNSKNPNVGDEKIDFKAPDPTDKELKEDIIVPAGSTSMGPRIDLLKPOVSAPGK 647
Db 564 -----ITLAMPN---OKYTEDKM--SDFTSYGVNLSFKRPDITAGG 602
Y 648 NIKSTLN 654
Db 603 NIMSTON 609

RESULT 2
P2P_LACPA STANDARD; PRT; 1902 AA.
ID P2P_LACPA 002470;
AC 002470;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-
DE ASSOCIATED SERINE PROTEINASE) (LP151).
DE PRP.
OS Lactobacillus paracasei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;

```

```

OC Lactobacillus.
OX NCBI_Taxid=1597;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 151;
RX MEDLINE=92381481; PubMed=1512565;
RA Holck A., Naes H.;
RT "Cloning, sequencing and expression of the gene encoding the cell-
RT envelope-associated proteinase from Lactobacillus paracasei subsp.
RT paracasei NCDO 151.";
RL J. Gen. Microbiol. 138:1353-1364(1992).
RN [2]
RP SEQUENCE OF 189-196.
RX MEDLINE=9226694; PubMed=1564442;
RA Naes H., Nissen-Meyer J.;
RT "Purification and N-terminal amino acid sequence determination of the
RT cell-wall-bound proteinase from Lactobacillus paracasei subsp.
RT paracasei.";
RL J. Gen. Microbiol. 138:313-318(1992).
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC SPECIFICITY, ALTHOUGH SOME SUBSTRATE PREFERENCE HAVE BEEN NOTED.
CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS, AND
CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC INSULIN B-CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M83946; AAA25248.1; -
DR PIR: B4858; B4858.
DR HSP: Q99405; IMPT.
DR MEROPS: S08.019; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR00137; PA.
DR InterPro: IPR00209; Peptidase_S8.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 3.
DR PRINTS: PRO0723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
DR HydroLase: Serine protease; Cell wall; Zymogen; Signal;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
SQ SEQUENCE 1902 AA; 200253 MW; D8C9F38CE5DA582 CRC64;

Query Match 12.4%; Score 429.5; DB 1; Length 1902;
Best Local Similarity 26.1%; Pred. No. 1e-13;
Matches 174; Conservative 97; Mismatches 275; Indels 121; Gaps 27;

```

QY	17	ISDKEAENKEKENHINSHAMETGQDKEKKTAIVKEKEVSNRPVNDNTNSMEAKIRE	76
Db	35	ISOQTKS-----SLANTYKAATKAQAQT--DTAATTNQAIAQIOLAAKGIDYK	82
QY	77	ENSKSGQDYDTSFVNKNTEPNKED---KVUYIAEFKDESEGEKAIKELSLKNTKVLV	133
Db	83	LNKVQOQDDTVYDVIVQNSAA-PASENGTLKTRDYSSTFAETIQEETNNKVIAAQSVAKEAVEQ	141
QY	134	T-----YDRIFNGSALETTPNMLDKIKQTEGSSVERQKQOPMNNNAKELIGEAL	186
Db	142	TQQTAGESYGVVNGVFSTKRVVNDIPRLKQIAGKVTTLAKVUYPTDAKANSMAVQAAM	201
QY	187	DYLKSIAPFEGKNDFNGMVFNSINDTIDVRHKAMRIDDAKASM--REKEDKLGTDKN-	244
Db	202	SNRY-----YKGEETVSVYDITDIDTPHKDMRLSDDKVKKLTIVYKEKTTQAKHG	252
QY	245	YWLSDKIPHAENVY--NGSKITVEKYDDGRDYFDRPHGMHAGILAGNDEQDINKFNGIDG	303
Db	253	RYFTSKVYGVGYNADNNDTIT---DDTVD--EQGHMNVAGIIGANCTGDD--PTKSVYG	304
QY	304	IAPNAQIFSYKMYSDAG--SGFAGDETFPHAIEDSIKNNVNVYSVSGF--TGTGLNGEKYV	361
Db	305	VAPFQOLLAMVVFNTNSDTSATTTGSAITVSAIEDSAKIGADVLNMSLGSDGNQLEDPET	364
QY	362	QAIRLAKRAGIPRVYVANGVNTSASSSMDLVANNHILKMTDTCVNTTAHEQIAIVASA	421
Db	365	AAVONANESGIAVAISAGNSSTSSATQG--YKNDYIGLDNDEKAVGTPGTSRCATTYASA	422
QY	422	KNQTFVEEDKVI-----GSESFYK--RNIGAFEDRSKITTNEDGTAKPSKLFVYIG	471
Db	423	ENTDVISOAVYITDQKDLQLPETIQLSNDFTSGFQKKFVYVKDASGDLK-----G	476
QY	472	KGQODDLGLDLRKRIVAMRITIKDLKNAFKK--AANDKARAIAMVNTVNYNTRDNMTE	529
Db	477	AAADYT---ADAKGRKIAIVKR--GELNFADKQRYAQAAGAAGLIYN-----NDGTA	523
QY	530	LPANGYEADECTKQVFSISGDDGVKL--NMNINPDKTKTEKVRNNKEDFKDLQEQYPID	587
Db	524	TPPLSIRLT--TTPTPTGLSSKTQCKLVDMYTAHPDLSGVK-----	503
QY	588	MESFNSNKPVNGDEKEIDFKFAPDPTDELYKEDIIVPAGSTSWGPRIDLKLKPDVAPGK	647
Db	584	-----IALTLLPN---QKYTEDKM--SDFSYGVYVNSLFSFKPDITAPGG	602
QY	648	NIKSTLN 654	
		603 NIWSTON 609	
RESULT 3			
P2P_LACLC			
P2P_LACLC	STANDARD:	PRT:	1902 AA.
AC	P15293;		
DT	01-APR-1990 (Rel. 14, Created)		
DT	01-APR-1990 (Rel. 14, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	PII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-ASSOCIATED SERINE PROTEINASE) (LP151).		
DE	ASSOCIATED SERINE PROTEINASE) (LP151).		
GN	PRT.		
OS	Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).		
OC	Plasmid pLP763.		
OC	Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC	Lactococcus		
OC	NCBI_TaxID=1359;		
OX	NCBI_TaxID=1359;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-NCDO 763;		
RA	MEDLINE=89313288; PubMed=2501630;		
RT	Kiwaki M., Ikemura H., Shimitzu-Kadota M., Hirashtina A.:		
RT	"Molecular characterization of a cell wall-associated proteinase gene		
RT	from Streptococcus lactis NCDO763."		
RL	Mol. Microbiol. 3:359-369(1989)		

[illegible]


```

CC -1- FUNCTION: THIS VIRULENCE FACTOR OF S.PYOGENES SPECIFICALLY CLEAVES
CC THE HUMAN SERUM CHEMOTAXIN CSA AT LYS(68)-ASP(69) BOND NEAR ITS
CC C-TERMINUS, DESTROYING ITS ABILITY TO SERVE AS A CHEMOATTRACTANT.
CC -1- SUBCELLULAR LOCATION: CELL WALL BOUND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J05229; AAA26960.1; -.
DR PIR: A35066; A35066.
DR HSP: P00782; 2SPT.
DR MEROPS: S08.020; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR003137; PA.
DR Pfam: PF00225; Gram_pos_anchor; 1.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 4.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
DR HydroLase: Serine protease; Signal; Cell wall; Repeat; Transmembrane.
FT SIGNAL 1 31
FT CHAIN 32 1167
FT DOMAIN 32 1139
FT TRANSMEM 1140 1157
FT DOMAIN 1158 1167
FT ACT_SITE 130 130
FT ACT_SITE 193 193
FT ACT_SITE 512 512
FT DOMAIN 1029 1104
FT DOMAIN 1034 1101
FT REPEAT 1034 1050
FT REPEAT 1051 1067
FT REPEAT 1068 1084
FT REPEAT 1085 1101
FT DOMAIN 1127 1131
FT PROTEINS.
FT SEQUENCE 1167 AA; 128263 MW; D2DDC32E5752DA5D CRC64;

Query Match 8.5%; Score 293; DB 1; Length 1167;
Best Local Similarity 23.0%; Pred. No. 2,7e-07;
Matches 133; Conservative 99; Mismatches 207; Indels 166; Gaps 29;

OY 97 NPKKEDKVVYIAEFKKEGSEKAI-----KELSLKNTKVLTYDRIFGNSAIEET 147
DB 24 NAQSDIKAMYTE--DPTVEQAVETPOPRAYSEVSPSSKRTKPPRPD-----DAETI 76
OY 148 PDNDKIKIOLIGISSEVERAKVVOPMNMHARKIEIVEEID--YAKSINAPGKPNDRGM 205
DB 77 ADDNDLAPAPAKTAD-----TPATSKATIRLDNDPSQVKTLOEKAGK--GAGT 124
OY 206 VINSIDTGTDRHAKMRIDDAKASMPFKEDLKGTDKNY-----WLSDKIPAFNYVN 259
DB 125 VVAVIDAGFDNHNHAWRLTDKTKARYO-SKEDLEKAKKEHITTYGGEWVNDKVAITYHDSK 183
OY 260 GCKITVEKYDDGRDYFDPHGMIAGILAGNTEODIKNFNGIDGIAFNAQIFSYKM---- 315
DB 184 DGKTAVDQ-----EHGTHVSGILSGN-APSETKERYPLEGAMPEAOQLLRVEIVN 233
OY 316 -YSAGSGFADDTMFHAIEDSIKHNVDVYSSGFGTGTGLVG--EKYQWQIRALRKAGI 372
DB 234 GLADYARNYA-----QAIRDAVNLGAKVINYMSFGNAALAVANLPDETKKAFDYAKSGV 287

```

```

OY 373 PMVATGCVATSSASSSSMDLVANNHLKMTDTCGNVTRTAHEDATAVA--SAKNOTVERDK 430
DB 288 SIVSAGSDSSFGGKTRPLA--DH--PDYGVGTTPAAASTILVASISPDOKLITERAM 342
OY 431 VNIGSESEKRYRNIGAFEDKSKITTNEGDTKAPSK-LKFVYIGKQDODLIGLDRGIAY 489
DB 343 VKTDQDQDK-----EMPLSTNR--FEPPNAYDYAVANRQMKEDDF-KDVKGKIAL 390
OY 490 MDRITTKOLKNAFKKAMDKGARALMVNTVNYNRD-----NMTELPAMGYEADGCT 541
DB 391 IER-GDIDFKKRVANAKKAGAVLI-----YDQDKGFPLELPVDMQPA----- 436
OY 542 KSQVFSISGDDGVKLMMNINPDKTEVRKNNKEDFKOLEYYPIDMESFNSKNPNVGE 601
DB 437 -----FISRKGLLLKD--NP-----Q 451
OY 602 KEIDFKFAPDPTDKELYEDITVPAGS-----TSWGPRIDLKPDVSAFGKNIKSTLN 654
DB 452 KTIFFNATPK-----VLPTASGTRKLSRFSWGLTADGNIKPIDAAPGODILS--S 499
OY 655 VINGK 659
DB 500 VANNK 504

RESULT 7
SCA2_STREPY STANDARD: PRT: 1161 AA.
ID SCA2_STREPY
AC P58099;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CSA PEPTIDASE PRECURSOR (EC 3.4.21.-) (Scp).
GN SCPA OR SPY2010.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yvan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus
RT pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
CC -1- FUNCTION: THIS VIRULENCE FACTOR OF S.PYOGENES SPECIFICALLY CLEAVES
CC THE HUMAN SERUM CHEMOTAXIN CSA AT LYS(68)-ASP(69) BOND NEAR ITS
CC C-TERMINUS, DESTROYING ITS ABILITY TO SERVE AS A CHEMOATTRACTANT.
CC -1- SUBCELLULAR LOCATION: CELL WALL BOUND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE006623; AK34691.1; -.
DR MEROPS: S08.020; -.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR HydroLase: Serine protease; Signal; Cell wall; Repeat; Transmembrane;
KW Complete proteome.
FT SIGNAL 1 31
FT BY SIMILARITY.

```

```

FT CHAIN 32 1181 C5A PEPTIDASE.
FT DOMAIN 32 1156 EXTRACELLULAR (POTENTIAL).
FT TRAMSEM 1157 1174 POTENTIAL.
FT DOMAIN 1175 1181 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 130 130 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 133 133 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 512 512 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1029 1121 CELL WALL ATTACHMENT DOMAIN (POTENTIAL).
FT DOMAIN 1034 1118 5 X 17 AA TANDEN REPEATS.
FT REPEAT 1034 1050 1.
FT REPEAT 1051 1067 2.
FT REPEAT 1068 1084 3.
FT REPEAT 1085 1101 4.
FT REPEAT 1102 1118 5.
FT DOMAIN 1144 1148 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
PROTEINS.
SQ SEQUENCE 1181 AA: 129486 MM: 3F9FC51763419CFC CRC64:

```

Query Match 8.5%; Score 292; DB 1; Length 1181;

Best local similarity 23.0%; Pred. No. 3.1e-07;

Matches 139; Conservative 93; Mismatches 207; Indels 166; Gaps 29;

```

QY 97 NPKKEDKVVYIAEFKDESGEKAI-----KELSLKNTKLVYTDRIENGSALETT 147
DB 24 NQSDIKANTVTE--DTPATEQAVETPOPTAVSEAPSSKETKTPQTPD-----DAEETI 76
QY 148 PDLKDKIKQIEGSSVERAKQVPMNMHAKKEIGVEALD--YLKSNAPFGKPFGRGM 205
DB 77 ADDANLAPQAPAKTAD-----TPATSKATYIRDLPDSQVKTLEKRGK--GAGT 124
QY 206 VISNIDTGTDRHAKARIDDDAKASMRFKEDLGTGDKNY-----WLSDKIPAFNYN 259
DB 125 VAVVIDAGEDKNHAKRLDRTKARYQ--SKEDLEKAKKEHGIYGEVNDKVAIYNDYSK 183
QY 260 GKKTYEKKDGDGNDYDPPHGMHITAGLAGNDTEODIKNFNGIDGAPNAOIESYKM---- 315
DB 184 DGTAVDQ-----EHGTHVSGILSGN--APSETKEPRLEGAPEQOLLMLRVEIVN 233
QY 316 -VSDAGSGAGDETMFHAIEDSIKHNVDVSVSSGFTGVLG--EKYQAIHALKRAKI 372
DB 234 GLADYARNTA-----QALIDAVNLCAKAYINMFGNNAALAYANLPETKKAEPYASKGV 287
QY 373 PMVATGNATYASASSSSWDLVANNHLKMTDGNVTRTAHEDAIAVA--SAKQVTEFDK 430
DB 288 SITVSAGNDSSEFGKTRPLA--DH---PDYGVVGTTPAADSTLYTAVSPDQLETTAT 342
QY 431 VNIGSEFYRNIGAFDPKSKITTNEDGTAKPSK-LKPYVIGKQDODLIGDLRGKIAV 489
DB 343 VKTADQODK-----EMKVLSTNR---FEPRKAYDAYAVANGKMEEDF-KDVKGKIAL 390
QY 490 MDRIYRKDKLNKFKKAMDGARIMVNVNYNRD-----NMTELPRAMQYEADGRT 541
DB 391 IER-GIDIRDKTAAKKAGAVGLT---YDNQKGFPEIPLNDQMPAA-----436
QY 542 KSGVFISGDDGVKLMNMIMPDKTEVKRNKKEDFDKLEQYYPIDMESEFNSKPNVGE 601
DB 437 -----FISKKDGLL-----KEN-----PQ 451
QY 602 KEIDFPAEDTKELYKEDIYVAGS-----TSMGPRIDLKLPDVSAPGNIKSTLN 654
DB 452 KTTTFNATRK-----VLPTASGTKLSRFSWGLTADGNIKPDIAAPGODILS--S 499
QY 655 VINGK 659
DB 500 VANNK 504

```

```

RESULT 8
SUBF_BACSU STANDARD: PRT: 1433 AA.
AC P16397;
DT 01-AUG-1990 (rel. 15, Created)

```

```

DT 01-FEB-1996 (rel. 33, last sequence update)
DE 20-AUG-2001 (rel. 40, last annotation update)
DT BACILOPEPTIDASE F PRECURSOR (EC 3.4.21.-) (ESTERASE) (RP-I PROTEASE)
DE (90 KDA SERINE PROTEINASE).
GN BPR OR BPF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 195-222.
RX MEDLINE=90170864; PubMed=2106512;
RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
RA Pero J.;
RT "Bacillopeptidase F of Bacillus subtilis: purification of the protein
RT and cloning of the gene.";
RL J. Bacteriol. 172:1470-1477(1990).
RN [2]
RP REVISIONS.
RC STRAIN=168;
RX MEDLINE=90368623; PubMed=2118514;
RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
RA Pero J.;
RL J. Bacteriol. 172:5520-5521(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90216713; PubMed=2108961;
RA Wu X.-C., Nathoo S., Pang A.S.-H., Carne T., Wang S.-L.;
RT "Cloning, genetic organization, and characterization of a structural
RT gene encoding bacillopeptidase F from Bacillus subtilis.";
RL J. Biol. Chem. 265:6845-6850(1990).
RN [4]
RP SEQUENCE OF 1-211 FROM N.A.
RX MEDLINE=89008180; PubMed=3139638;
RA Beall B., Lowe M., Lutkenhaus J.;
RT "Cloning and characterization of Bacillus subtilis homologs of
RT Escherichia coli cell division genesftsZ andftsA.";
RL J. Bacteriol. 170:4855-4864(1988).
RN [5]
RP SEQUENCE OF 1410-1433 FROM N.A.
RC STRAIN=168 / MARBURG;
RX MEDLINE=90174995; PubMed=2106671;
RA Masuda E.S., Anaguchi H., Sato T., Takeuchi M., Kobayashi Y.;
RT "Nucleotide sequence of the sporulation gene spoIIIGA from Bacillus
RT subtilis.";
RL Nucleic Acids Res. 18:657-657(1990).
RN [6]
RP SEQUENCE OF 195-219.
RC STRAIN=NATTO 16;
RA Kato T., Yamagata Y., Arai T., Ichishima E.;
RT "Purification of a new extracellular 90-kDa serine proteinase with
RT isoelectric point of 3.9 from Bacillus subtilis (natto) and
RT elucidation of its distinct mode of action.";
RL Biosci. Biotechnol. Biochem. 56:1166-1168(1992).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sb.ch/announce/
CC or send an email to license@isb-sb.ch).
CC -----
CC EMBL: M29035; AAA62679.1; -
CC EMBL: J05400; AAA83362.1; -
CC EMBL: M22630; AAA22458.1; -
CC EMBL: X17344; CAA35224.1; -
CC EMBL: Z89111; CAB13403.1; -
CC EMBL: Z89112; CAB13404.1; -
CC PIR: A35131; A35131.

```

DR PIR: A35750; A35750.
 DR PIR: A36734; A36734.
 DR HSP: P00782; 2SBT.
 DR MEROPS: S08_017; -.
 DR SUBCLIST: Bg10233; bpt.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 DR HydroLase: Serine protease; Zymogen; Signal; Complete proteome.
 KW SIGNAL 1
 FT PROPEP 31 194
 FT CHAIN 195 755
 FT PROPEP 756 1433
 FT ACT_SITE 227 227
 FT ACT_SITE 274 274
 FT ACT_SITE 452 452
 FT CONFLICT 219 219
 FT CONFLICT 393 393
 FT CONFLICT 829 834
 FT CONFLICT 836 841
 FT CONFLICT 844 852
 FT CONFLICT 853 1433
 SQ SEQUENCE 1433 AA; 154577 MW; 98DF6846897807C9 CRC64;

Query Match 6.98; Score 239.5; DB 1; Length 1433;
 Best Local Similarity 20.58; Pred. No. 0.00015;
 Matches 156; Conservative 128; Mismatches 257; Indels 221; Gaps 37;

10 SDEALNTSDKEAENKEHENHISAETQDFEKKTVIKKEVYKRPVIDNNTSN 69
 30 ASSKVTGSYVKEKQSAESIONKISSSLKSKFKKEKTEFLIKRKDLA--NPEAKAAAY 87
 70 EEARKEENSKSOGDYTFVFN--KNTENPKKEDKVYIAEFKDESEGAKEIKELSSLK 127
 88 KKAKSKLSAKTEKQSAVSSILKYADESQDYLKYLNTQDKGNADQ----- 138
 128 NTKVLTYDRFNGSAIETTDPNDIKIQIGISYEAAQVQ-----PM-MHARKEI 180
 139 ----IHST-YVANGIAVHASKVEKVKVQFEVEKVLNPEKROLEKSSSPNMKKAOKAI 193
 181 GVEAIDY-LKSIAP--FGKFDGRGVINIDTGTFRKAKRIDDAKMRKKED 237
 194 KATGGEVNNVQIDAPKAMALGDTGTGVVASIDTGVMMNPAL-----KEK 240
 238 LKGTDKNWLSDKIPHAENYNGKITYEDGRDYFDPGHMIIAGILAGNDTEQDIKN 297
 241 YRGYNPEN--PNEPENENMYDAVAGEASPYD-----LAHGTHTGTMGSEPD---G 289
 298 FNGIDGIAPNQIFSYKMYSDAGSG-----FAGDETMAHLEDISKH--NVDVVSSSG 349
 290 TNOI-GVAPGAKMTAVAKFSEDSGTDADILEAGEVTLAPKDAEGNPHPEMAPDVANNWG 348
 350 FTGTGLGGEKQWQAIRLAKGIPMVATGN-----YATSSA-- 386
 349 -GGSGL-DEWYRDVNMNRADIFEPFSAGNTDLFIPGPGSIANPNANPESPATGADI 406
 387 -----SSWD--LVANNHLK---MTD 402
 407 NKKLADESLQSPSYDEIKPEISAPGVNIRSSVPQGYEDMDGTSMAGPHVASVAALLK 466
 403 TGNVTRFAHEDAIJAVASAKNQYEE--DKVNIAGESKFKYKNIGAFPKSKITINEDCT 459
 467 QANASLSVDEKEDILTSTAEPSTPDSPNNG---YGHVLNAPFAVASAVT---DELG 520
 460 KAPKELFVYIGKQDOD-----LIGDLRCKIIVMVRITYTKDKNAFKKMD 507
 521 KAEQVSV-----EGDDDEPPYQHEKYTEAELEGSLPITLAEDNVSTYTKLSYK--LD 574
 508 KGARAIWVNTVNYNRDNMTLPAKMGVEADGTRKSOVFSISGDDGVKLMNMNPDKKTE 567

DB 575 QG-----EMTEIRAK-----RISCHLAKCTYQAEIPDIK-- 603
 QY 568 VKRNKKEDEFKKLEQY--YPIDMESFN-SNKPNNQDEKEDFKRPAPODRKLEKEDIYVP 624
 DB 604 --GTRKSKWMIHDEFGHVSSDYDYTVKPSITAGYKQDFEAPAG-----WVA 651
 QY 625 AGSTS--WGPRIIDLKPDVASAPCKNIKSTLVNNGKSTYVG 663
 DB 652 SCTNNNMWNG-----VPSGCPN-----TASGEKXYVG 678

RESULT 9
 WPIA_BACSU STANDARD; PRT: 894 AA.
 AC P54423; 006726;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CELL WALL-ASSOCIATED PROTEASE PRECURSOR (EC 3.4.21.-) [CONTAINS: CELL WALL-ASSOCIATED POLYPEPTIDES CWBP23 AND CWBP52].
 GN WPIA.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_Taxid=1423;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-54 AND 414-428.
 RC STRAIN=168;
 RX MEDLINE=97158234; PubMed=9004506;
 RA Margot P., Karamata D.;
 RT "The wprA gene of Bacillus subtilis 168, expressed during exponential growth, encodes a cell-wall-associated protease.";
 RT Microbiology 142:3437-3444(1996).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=98015415; PubMed=9353931;
 RX Medina N., Vanlier F., Roche B., Autret S., Levine A., Seror S.J.;
 RA "Sequencing of regions downstream of addA (98 degrees) and cItg (289 degrees) in Bacillus subtilis.";
 RT Microbiology 143:3305-3308(1997).
 CC -1- FUNCTION: NOT YET KNOWN; COULD BE INVOLVED IN PROTEOLYTICAN
 CC DEGRADATION, BY CLEAVAGE OF ITS PEPTIDE BRIDERS.
 CC -1- SUBCELLULAR LOCATION: CELL-WALL BOUND.
 CC -1- PFM: PROCESSED INTO CWBP23 AND CWBP52.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U58981; AAC25926.1; -
 DR EMBL: Y09476; CAAT0641.1; -
 DR EMBL: 299109; CAB12917.1; -
 DR HSP: Q99405; 1MPT.
 DR MEROPS: S08_004; -
 DR Subtilist: Bg11846; WPIA.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KW HydroLase: Serine protease; Cell wall; Zymogen; Signal;
 KW Complete proteome.
 FT SIGNAL 1 31
 FT CHAIN 32 894
 FT CHAIN 894 ?
 FT CELL WALL-ASSOCIATED PROTEASE.
 FT CWBP23.


```

FT PROPEP 413 POTENTIAL.
FT CHAIN 414 CMBP52.
FT ACT_SITE 466 RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 497 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 650 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CONFLICT 9 V -> A (TN REF. 1).
FT CONFLICT 14 L -> I (TN REF. 1).
SQ SEQUENCE 894 AA: 96487 MW: 0F67C353E5FBD8C CRC64;

Query Match 5.8%; Score 201.5; DB 1; Length 894;
Best Local Similarity 19.8%; Pred. No. 0.0059;
Matches 155; Conservative 112; Mismatches 247; Indels 269; Gaps 40;

OY 5 LADTSSSEDALNISDK-----EKVAENKEKHNIHSAMETSO-----DFKE---KRTAVI 51
DB 200 LSGTAGSKELTSLYKAAPIYAKLALNKARREIYODLTLPLEDVSENGASSSYKV 259
OY 52 KEKEVYVSKNPVIDNNTSNEBAKIEE-----NSKSGDYTDSFVNKNTENPKKE-- 101
DB 260 TEKQKAKINRLYKALQVSPFLKEEIKQADRLNMKQLOGKTAGAILTENNTIAKSEVQ 319
OY 102 -DKVYVIA-----EFKDESEKAIKELSLKNTKVLTYDRIF----- 139
DB 320 YTVVIRKVKDNKSLSSVHNEMKGFSAQSKDISVYKAKKL--FDNLXSFELPDEKQ 377
OY 140 NGS-----AIETTPNLDKIKQIEGISSYER---AQKQO-----PMNHARKETGVEE 184
DB 378 NGVYTSASAKRVKAAATLMSMSNVEFAEPYQEKSLANDQYQYQWPLKNGENG--GVKN 436
OY 185 AIDYKLSINAPFEGKNFDRGCMVISNIDTGTDYHKKAMRIDDAKASMRKEDLKG--T 241
DB 437 ADVKYPEPANTLTSKR--KLNDTLIAVDTGVD-----STLADLKGVRT 478
OY 242 DKVYWLSDKIPHAFNYYNGKIVYKIDGRODYFDPGHMIIAGILAGNDQEDIKNNGI 301
DB 479 D-----LGHNFVGRNNAMD-----DOG-----HGHVAGILTAQS-----DNQYSM 515
OY 302 DGIAPNAQIFSYKMYSDAGSG-----FAGDE-----TMFHAIEDSIKHN 340
DB 516 TGLNAKAKIIPVAVLDSAGSGDTEQALGIKYAKADKAKYINLSLGGYSRVLEFALKYA 575
OY 341 VD---VVSVSSTGTGGLV---GEKYQWAIIRLR-----AGIP 373
DB 576 ADKNVILAAASGDGENALSYPASSKYVSGATNRMDTADFSNYGKGLDISAPGSDIP 635
OY 374 MYVATGN-----YATSSASSMDLVANNHKKMTDGNVTRTAHEDAIANASA 421
DB 636 SLVPNGNVYTMSTGMTAPYAAAAGLLF--AONPKLRTEVEDMKKTADD----- 685
OY 422 KNOTVEFDKVNIGESFVKYRNIGAFDPSKITTNEDGTAPSKLKFVYIGKODDOLIGL 481
DB 686 -----ISFESVD--GGEELIYDYG---DPIEI-----PKPI-----GV 714
OY 482 DLR---GKIAVMDRIYTKDLKNAFKAMDKGARAIVNVTNVRNDRNTELPAMGYEAD 538
DB 715 DMHSGYGLRLVMAKVASADLQ-----LKVNLKES-----TQTAVRG--SAK 753
OY 539 EGRKSOVFSISGDDGVKLMNMIPDKTEYKRNKKEDFKLEQIYPIIDMESNSKNPNV 598
DB 754 EGT-----LIEVNMNGKRLKLSAKAGKD-----NAFVNI 782
OY 599 GDKE---IDFKFAPDTPDEKLEYEDIIIVPAGSTSWGPRIDLKLPDVSAPGK--NIKSTLN 654
DB 783 ATQKQOVLYLKATKGDAKTSYK--VYVYVKGKSGTGPVNAVAKTKDTAVAGKANSKAMIR 840
OY 655 VIN 657
DB 841 YKN 843

```

```

ID PRTB_YEAST STANDARD: PRT: 635 AA.
AC P09232:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE CEREYISIN PRECURSOR (EC 3.4.21.48) (VACUOLAR PROTEASE B) (PROTEINASE
DE YSCB).
GN PRB1 OR YEL060C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 281-295.
RC STRAIN=AB320;
RX MEDLINE=88142830; PubMed=3325823;
RA Moehle C.M., Rizard R., Lemmon S.K., Smart J., Jones E.W.;
RT "Protease B of the lysosomal vacuole of the yeast Saccharomyces
RT cerevisiae is homologous to the subtilisin family of serine
RT proteases."
RL Mol. Cell. Biol. 7:4390-4399(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunkeler-Smith S.,
RA Hyman R., Kayser A., Komp C., Lasker D., Lew H., Lin D.,
RA Moseedale D., Nakahara K., Namath A., Norgren R., Oelner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yellon M., Botstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 605-635 FROM N.A.
RC STRAIN=S288C;
RA Saunders W.S., He L., Loo K.K., Hoyt M.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP PROCESSING.
RX MEDLINE=92078140; PubMed=1744078;
RA Nebes V.L., Jones E.W.;
RT "Activation of the proteinase B precursor of the yeast Saccharomyces
RT cerevisiae by autocatalysis and by an internal sequence."
RT J. Biol. Chem. 266:22851-22857(1991).
CC -1- FUNCTION: AMONG OTHER SUBSTRATES, ACT ON YSCA (PEP4) TO ACTIVATE
CC IT BY PROCESSING ITS PRO-PEPTIDE.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS WITH BROAD SPECIFICITY,
CC AND OF BZ-ARG-OET > AC-TYR-OET. DOES NOT HYDROLYSE PEPTIDE AMIDES.
CC -1- SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M18097; AAA34901.1; -
CC EMBL: U18795; AAB65027.1; -
CC EMBL: Z11859; CAA77886.1; -
CC EMBL: M80522; AAA34495.1; -
CC PIR: A29358; A29358.
CC HSSP: P06873; 3PRK.
CC MEROPS: S08.052; -.
CC SGD: S0000786; PRB1.
CC InterPro: IPR000209; Peptidase_S8.
CC Pfam: PF00082; Peptidase_S8; 1.
CC PRINTS: PR00723; SUBTILISIN.
CC PROSITE: PS00136; SUBTILASE_ASP; 1.
CC PROSITE: PS00137; SUBTILASE_HIS; 1.

```

RESULT 10
PRTB_YEAST

DR PROSITE: PS00138: SUBTILASE_S8; 1.
 KW Hydrolyase; Serine protease; Glycoprotein; zymogen; signal.
 FT SIGNAL 19
 FT PROPEP 20 280
 FT CHAIN 281 635
 FT DISULFID 460 491
 FT ACT_SITE 325 325
 FT ACT_SITE 357 357
 FT ACT_SITE 519 519
 FT CARBOHYD 594 594
 FT CONFLICT 622 622
 SQ SEQUENCE 635 AA; 69621 MW; C346C2B1C7DDDC48 CRC64;

Query Match 5.68; Score 194; DB 1; Length 635;
 Best Local Similarity 19.18; Pred. No. 0.0088;
 Matches 121; Conservative 90; Mismatches 207; Indels 214; Gaps 27;

QY 21 EKVAENKREHINISAMETSDQFEKKTAIVKEKESKPNVIDNNTSNEAKIKESNSN 80
 b 149 EKTLEKHHNRRLAPLVSTAQPNDAISKITPNRIIT---YFKGAPQGEIDPFKEVQ 204
 QY 81 KSGQDYDSEVKNKTENPKKEDKVVYIAEFKDESGEKAELSLKNTKVLTYTDRIEN 140
 Db 205 QAO-----LQSEVNLSEAD-----AFISTKDTSLSTSEAGIIDS---FNINDLFS 248
 QY 141 GSAIETTPDNLDKIKOIGISSVERAKVQPMNNHARKEIVEEAIDYLSINAPFG--- 197
 Db 249 GYIGFTYEIDLRKQNPFLVPEVD-----SIVEETEDTONSAWGLAR 294
 QY 198 -----KNFD-----GRGVISNIDTGYDHRKAMRIDDAKA---SMRFRKE 236
 Db 295 ISHREMLNLSGFNKLVDYDDAGRGVTSYIDTVGINIK---DEKKAINGKITPLNDE 350
 QY 237 DLKGTDKMYLSDKIPAFNNTNGSKITVEKYDDGRYDFDPHGMHIALGNTDQDIK 296
 Db 351 DLDG-----NG-----HGTICAGTIAS-----K 368
 QY 297 NFNIGIDTAPNAOTFSYKMYSDAGSGFAGDETMFHAIDSIK-HNVDVSVSSGPTGT-- 353
 Db 369 HY-----GVAKANNAVAVVLSNNGSGTMSD--VVKGVYIAAKAHOKKEKKKFGKSTA 422
 QY 354 --GLVGEK--YMOAIRALKRAGIPMVAVATSNATSSMDVLANNHLMKMTDGNVTR 408
 Db 423 NMSLGGGKSPALDLAVNAVEGIFHAAGNENDACNTS----- 463
 QY 409 TAAHEDATAVASAKNQIYEPDK-----VNIGSESKYRNIGAFPDKSKI 452
 Db 464 PASADKALTVGAS---TLSDRAVFSNMKGVDFAPGLNI-----LSTYIGSDA 511
 QY 453 TTNEGDTKAPSK-----LKFVYIGKGODDGLGLRGKIAVMDRITYTKDLKNAFKAM 506
 Db 512 TATLSGTSMASPHVAGLITFLSLQPGSDSEFFELG-----QDSLTPQDLK 558
 QY 507 DKGARAIWVTVNYNBDNNTTELPRMGVEADEGTRSQVFSISGDDGVKLMNMINPDKKT 566
 Db 559 -----KLHISTKDILEDIP-----EDTPNVLIYNGGDLAFNWN----- 594
 QY 567 EVKRNKDEPKDLEQYPIFIMESNSKNPNV 598
 Db 595 DTKRSHSGFKQEL-----NMDEFISGKIDL 620

RESULT 11
 THES_BACSP STANDARD: PRT: 401 AA.
 AC 045670;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE THERMOPHILIC SERINE PROTEINASE PRECURSOR (EC 3.4.21.-) (AK.1
 DE PROTEASE).
 OS Bacillus sp. (strain AK1).

OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OC NCBI_TaxID-1409;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE-95085262; PubMed-7993087;
 RA Maciver B., Mchale R.H., Saul D.J., Bergquist P.L.;
 RT "Cloning and sequencing of a serine protease gene from a
 thermophilic Bacillus species and its expression in Escherichia
 coli";
 RL Appl. Environ. Microbiol. 60:3981-3988(1994).
 RN (2)
 RP X-RAY CRYSTALLOGRAPHY (1.80 ANGSTROMS) OF 122-401.
 RX MEDLINE-20057863; PubMed-10588904;
 RA Smith C.A., Toogood H.S., Baker H.M., Daniel R.M., Baker E.N.;
 RT "Calcium-mediated thermostability in the subtilisin superfamily: the
 crystal structure of Bacillus Ak.1 protease at 1.8-A resolution";
 RL J. Mol. Biol. 294:1027-1040(1999).
 CC -1- COFACTOR: CALCIUM-DEPENDENT. BINDS 4 CALCIUM IONS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- MISCELLANEOUS: HAS A PH OPTIMUM OF 8.5, A TEMPERATURE OPTIMUM OF
 75 DEGREES CELSIUS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 SUBTILASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L29506; AAA63688.1; -;
 CC PDB: 1DBT; 18-NOV-99.
 DR MEROPS: S08.009; -;
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS: PRO0723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KW Hydrolyase; Serine protease; zymogen; signal; calcium-binding;
 KW 3D-structure.
 FT SIGNAL 1 24
 FT PROPEP 25 121
 FT CHAIN 122 401
 FT ACT_SITE 160 160
 FT ACT_SITE 193 193
 FT ACT_SITE 347 347
 FT DISULFID 258 260
 SQ SEQUENCE 401 AA; 42835 MW; 1C736EFA489F256F CRC64;

Query Match 5.58; Score 191; DB 1; Length 401;
 Best Local Similarity 21.75; Pred. No. 0.0068;
 Matches 107; Conservative 61; Mismatches 130; Indels 196; Gaps 23;

QY 51 IKEREVNSKN-----PVINDNNTSNEAK-----IKENSNKSGQGTDSFVKNKNTNP 98
 Db 1 MKFAIVSLAVSWSLPPFLVEAASNDGVSPKTVSEINVSHEKGAVQGEV----- 53
 QY 99 KKEDKVVYIAEFKDESGEKAIKELSLKNTKVLTYTDRIENFGSAIETTPNLDKIK--- 155
 Db 54 -----IYQFKEDVNAEKAALKEVGATAV-----PDN-DRVKEKF 88
 QY 156 --QIEGISSVERAKVQPMNNHARKEIGVEEAIDYLSINAPFGKN---FDG----- 202
 Db 89 NVLKVGNVEAVKALNNPNLVEYAEP-----NYL--FNAAMPNDTYGYGYGQPN 138
 QY 203 -----RGMVSNIDTGYDHRKAMRIDDAKASMRKEDLKTGDKMYLWLD 249
 Db 139 TYTDYAMDVTKSGSGQELAVIDTGVDTYH-----PDLGK----- 173

```

QY 250 KIPAEYNGKKTVEKDYDGRDYP-----HGHIAGLACNDTEODIKNFNGIDGI 304
D 174 -----VIKGYFVDNDYDPMDLNHNHGTAVAGIAA-----ETNNATGIAGM 214
QY 305 APNAQISYKMYSDAGSGFADETMFHAIEDSIKHNVD-----VYSVSSG---FTGTGYG 357
D 215 APNTRILAVRLDRNGSGTSLD-----IADAILYADSGAEVNLISLGCDCCHTTT----- 264
QY 358 EKYQAIRALAKAGIPMVVATGATNATY-----ASSSSW----- 390
D 265 --LENAVYVAMNKSYYVAAAGNNGSSSTTEPASTENVIANGAVDYDRLASFSGYVW 322
QY 391 DLVA-----NNHLKMTDGTGNTRTAHEDAIAVASAKN-----QTVF--FDKVN 432
D 323 DVNAPGVDIYSTIGTGNRYAVSGTSMASPHVAGLALLASGRNNIEIRQAIETQADKIS 382
QY 433 IGSEFMYRNGAF 446
D 383 GTGYFMYRNGRINSY 396

RESULT 12
SUBN_BACNA STANDARD; PRT; 381 AA.
AC P35835;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SUBTILISIN NAT PRECURSOR (EC 3.4.21.62).
GN APRN.
OS Bacillus subtilis var. natto.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86029;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NC2-1;
RX MEDLINE-93113095; PubMed-1369081;
RA Nakamura T., Yamagata Y., Ichishima E.;
RT "Nucleotide sequence of the subtilisin NAT gene, aprn, of Bacillus
RT subtilis (natto).";
RL Biosci. Biotechnol. Blochem. 56:1869-1871(1992).
CC -1- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS WITH BROAD SPECIFICITY
CC FOR PEPTIDE BONDS, AND A PREFERENCE FOR A LARGE UNCHARGED RESIDUE
CC IN P1. HYDROLYSES PEPTIDE AMIDES.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
CC SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY
CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
CC IS NOT NECESSARY FOR NORMAL SPOULATION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D25319; BAA04989.1; -
DR EMBL: S51909; AAC60424.1; -
DR PIR: JH0778; JH0778.
DR HSP: P07518; IME.
DR MEROPS: S08.001; -
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.

```

```

DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Hydrolyase; Sporulation; Serine protease; zymogen; signal.
FT SIGNAL 1 23
FT PROPEP 24 106
FT CHAIN 107 381
FT ACT_SITE 138 138
FT ACT_SITE 170 170
FT ACT_SITE 327 327
FT ACT_SITE 381 AA; 39507 MW; DAEDABIEDIBAO92 CRC64;
SQ SEQUENCE

Query Match 5.58; Score 188.5; DB 1; Length 381;
Best Local Similarly 22.58; Pred. No. 0.0084;
Matches 80; Conservative 61; Mismatches 131; Indels 83; Gaps 15;

QY 86 YTDSFVKNKTEENP-KKEDKVYVIAEPKDESGEKA--KELSLKNTKVLTYDRIENG 142
D 18 FTMKFSNMSQAAGKSSSTTEKRYIVGFQTMASMSAKKDDISCKGVQKQF-KYVMAA 76
QY 143 ALETPPNLDKIKQIEGSSYERAKQVPMNHRKEIGVEAIDY-LKSINAP--FGKN 199
D 77 AATIDEKAVKELKKDPYAYVE-----DHIAHEVA--QSVPGISQIKAPALHSQG 126
QY 200 FDGRGMYISNIDTGYDHRKAMRIDDDAKASMRKKEDLAKTDKNYLSDKIPAEFY 259
D 127 YTGSNVKAVIDSGIDSSHPLNVRGASF-----VPEETNPYQ 165
QY 260 GCKTVEYDGGGRYFDPHGHIAGLACNDTEODIKNFNGIDGIAPNAQISYKMYSDA 319
D 166 DG-----SSHGTHVAGTIA-----LNNSIGVLVAASASLYAKVLDST 205
QY 320 GSGFAGDETFHAIEDSIKHNVDVSVS-SGFTGLGKGYMOAIRAKRAGIPMVVAT 378
D 206 GSG--QYSWIINGEMALSNMMDYINMSLGGPTSTALKTYVDAAV-----SSGIVVAAA 259
QY 379 GNYATSSASSMDDLVAANHLMKMTDGTGNTRTAHEDAIAV--ASAKNOTVEFDKV 431
D 260 GNESSGSGSTP-----VGYPAKYPSTIAGVAVSSNQRASSFSSV 298

RESULT 13
SUBE_BACSU STANDARD; PRT; 645 AA.
AC P16396;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MINOR EXTRACELLULAR PROTEASE EPR PRECURSOR (EC 3.4.21.-).
GN EPR OR IPA-15R.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / DB204;
RX MEDLINE-90340301; PubMed-2116590;
RA Brueckner R., Shoseyov O., Dol R.H.;
RT "Multiple active forms of a novel serine protease from Bacillus
RT subtilis.";
RL Mol. Gen. Genet. 221:486-490(1990).
CC [2]
CC SEQUENCE FROM N.A.
RP MEDLINE-89053875; PubMed-3142851;
RX SIOMA A., Ally A., Ally D., Pero J.;
RT "Gene encoding a minor extracellular protease in Bacillus subtilis.";
RL J. Bacteriol. 170:5557-5563(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE-95020537; PubMed-7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,

```

RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
 RA Pressecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
 RA Rapoport G., Danchin A.;
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97
 RT kb region from 325 degrees to 333 degrees."
 RL Mol. Microbiol. 10:371-384(1993).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X53307; CA37392.1; -
 DR EMBL; M22407; AAA22423.1; -
 DR EMBL; X73124; CA51571.1; -
 DR EMBL; 299123; CA51566.1; -
 DR PIR; S11504; SUBSMP.
 DR PIR; S39670; S39670.
 DR HSSP; 099405; 1MPT.
 DR Subtilist; BG10561; epr.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILASIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolyase; Serine protease; zymogen; Signal; Complete proteome.
 FT SIGNAL 1 27 POTENTIAL.
 FT PROPEP 28 103 POTENTIAL.
 FT CHAIN 104 645 MINOR EXTRACELLULAR PROTEINASE EPR.
 FT ACT_SITE 142 142 CHANGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 172 172 CHANGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 326 326 CHANGE RELAY SYSTEM (BY SIMILARITY).
 SO SEQUENCE 645 AA; 69695 MW; 401AD5B60BE2EA CRC64;

Query Match 5.48; Score 187; DB 1; Length 645;
 Best Local Similarity 22.1%; Pred. No. 0.02;
 Matches 136; Conservative 74; Mismatches 191; Indels 214; Gaps 33;

QY 175 HARRIGVEEADIDYLSKSNAPKGFNDRGMYI-SNIDTGTYYRH-KAMRIDDKAKSMR 232
 DB 26 HAQNSSEKEVIVY-----KMKAGKETLDSADVEQYKHLPAVAATVDETYKE 76
 QY 233 FKKE-----DLK-----GTD-----KNYV---LSDK-- 250
 DB 77 LKQPDILYVENNVSTAASTDFKVLSDGDTSDNFQEMNLEPTQVQKAMKAGLTGKRI 136
 QY 251 -----IPAFNYNGKRTVE-----KYDDGDFYDPFGMHIGAGILAGNDTEODIK 296
 DB 137 KIAVIDSGISPHDLSIAGXSAVSYTSSYKDDNG-----HGTAVAGILGAK-----H 184
 QY 297 NFNQIDGIAIPNAQITSYKMSYDAGSGFAGDETFMFAIEDSKIHNDVYSVSGFGTGLV 356
 DB 185 NGYIGDGIAPPAQIYAAVALDONGSGL--OSLLOGIDMSIANRDMIVMSLGTTSDSKI 242
 QY 357 GEKYWQAIIRALKKAGIPWVATGNAVTSASSSSMDLVANNHMKMDTGN---VTPTAAHE 413
 DB 243 ---LHDVAVKAYEGCVLLVAASGN-----DGKAKVNTPAAS 277
 QY 414 DAIVASA--KNQVEFDKVNIGESFXY-----RNIGAFEDKSKITTEDGTAKA-PSKIK 466
 DB 278 SVAVASATNENKQNLASFST---GQEVFSAAGTNTSYLYNQYATSGTSGQARPHAAA 334
 QY 467 FVYIGKGDODDGLIGDLGKIAVMDRIYTKDKNAFKRAMDKGARAINVYNTVNTYNRDN 526
 DB 335 MFALLKQDPAPETNVLREEMR-----KNIVDGLTA-----GRDQ 369

QY 527 WTELPAWGVEAD-----EGTKSQVESISGDDGVKILNMI-----NPDKKT 566
 DB 370 QREYGLIYKKAQITSDATYAAAEQAVKKABOTKNOI-----DLNKARELISQLPNSDAKT 423
 QY 567 EVKR-----NKKEDFKDKL--EYQ---YVIDMESFNSK-PVNGDEKE-----603
 DB 424 ALHKRLDKVQSYRNVDADKVAKEKRYTQGTVDTAQTAINKLPNGTOKKNLQKRLDQV 483
 QY 604 ---IDKFPAPD---TDKELYKEDI---IYPAGS--TSMGPRIDLLKPDVSAPG 646
 DB 484 KRYIASKQAKDKYAKAKSKKRTVDVSQASIGKLPASSSEKTSLOKRLNKVSKTNL----539
 QY 647 KNKSTLVNINCKST 661
 DB 540 KTAQGSVAAEKKST 554

RESULT 14
 TAGB_DICDI
 ID TAGB_DICDI STANDARD; PRT: 1905 AA.
 AC P54683;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 20-AUG-2001 (rel. 40, Last annotation update)
 DE PRESTALK-SPECIFIC PROTEIN TAGB PRECURSOR (EC 3.4.21.-).
 GN TAGB.
 OS Dictyostellium discoideum (Slime mold).
 OC Eukaryota; Eukaryota; Dictyostellida; Dictyostellium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RX MEDLINE=955262903; PubMed=7744252;
 RA Shaulsky G., Kuspa A., Loomis W.F.;
 RT "A multidrug resistance transporter/serine protease gene is required
 RT for prestalk specialization in Dictyostellium."
 RL Genes Dev. 9:1111-1122(1995).
 CC -1- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDIATE
 CC INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY
 CC S8; ALSO KNOWN AS THE SUBTILASE FAMILY.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
 CC TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS), MDR SUBFAMILY.
 CC -1- SIMILARITY: STRONG, TO TAGC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; U20432; AAA62212.1; -
 DR HSSP; P13569; INBD.
 DR Dictydb; DD02059; tagb.
 DR InterPro; IPR003593; AAA.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR001140; ABC_transporter_tmam.
 DR InterPro; IPR001687; ATP_GTP_A.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00664; ABC_membrane; 1.
 DR Pfam; PF00005; ABC_tran; 1.
 DR PRINTS; PR00723; SUBTILASIN.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW Hydrolyase; Serine protease; ATP-binding; Transport; Transmembrane;
 KW SIGNAL 1 ? POTENTIAL.

SUBTILASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D00264; BAA00186.1; -
DR PIR: A00971; SUBSS.
DR PIR: A41448; A41448.
DR HSSP: P07518; IMEE.
DR MEROPS: S08.001; -
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Hydrolyase; Sporulation; Serine protease; Zymogen; Signal.
FT SIGNAL 1 30
FT PROPEP 31 106
FT CHAIN 107 381
FT ACT_SITE 138 138
FT ACT_SITE 170 170
FT ACT_SITE 327 327
FT CONFLICT 191 191
FT CONFLICT 365 365
SQ SEQUENCE 381 AA; 39467 MW; 2251BADE22B4824F CRC64;
POTENTIAL.
SUBTILISIN AMYLOSACCHARITICUS.
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
S -> A (IN REF. 2).
N -> D (IN REF. 2).

Query Match 5.4%; Score 185.5; DB 1; Length 381;
Best Local Similarity 22.8%; Pred. No. 0.012;

Matches 81; Conservative 59; Mismatches 126; Indels 89; Gaps 16;

QY 86 YTDSEFNKNTENP-KKEKDVYIAEFKDESGEKAI--KELSLKNTKVLTYDRIFNGS 142
DB 18 FTMAFSNMSAQAAGKSSTEKKYIVGFQOTMSAMSAKKKDYISEKGKVKQKF-KYVNAA 76
QY 143 ALETPDNLKIKOIEGISSVERAKQVOPMMNHAKKEIGVEAIDY-LKSINAP--FGKN 199
DB 77 AATLDEKAVKELKKDPVAAYVEE-----DHIAHEYA--QSVPYGISQIKAPALHSQG 126
QY 200 FDGGMVINSNDTGTDRYHKKAMRIDDOAKAMREKKEDLKGTDKNYWLSDKIPHAENYYN 259
DB 127 YTGSNVAVYIDSGIDSSHPLNVRGASF-----VPSETNPYQ 165
QY 260 CGKITVEKYDDGRDYFDPGHMHIAGILAGNDTEODIKNFNGIDGIAPNAQIFSYKMYSDA 319
DB 166 DG-----SSHGTHVAGTIAA-----LNNSIGLVGSPSASLXAVKVLDDST 205
QY 320 GSGFAGDETFHAIEDSTKHNVDVSVS---SGFTGTGLVGEKRYWQAIRALRKAGIPMV 375
DB 206 GSG--QYSMIINGIEMWISNMNDIYNNISLGGPSCGSLTKTYVDK--AV---SSGIYVA 256
QY 376 VATGNVYATSASSSSMDLVANNHLMKMTDGNTRTAHEDAIAV--ASAKNOTVEF 428
DB 257 AAAGNEGSSGSSST-----VGYPAKYPTSTIAVAGAVNSSNQRASF 295

Search completed: December 7, 2001, 11:49:51
Job time: 205 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2001, 11:46:10 ; Search time 28.64 Seconds
(without alignments)
3391.230 Million cell updates/sec

Title: US-09-590-991-6
3454
Perfect score: 1 YPVVLADTSSSEDALNISDK.....PEKNIKSTLVNKGKSTYGV 664
Sequence:

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
tal number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3454	100.0	2119	2 Q9AHT5	Q9AHT5 streptococc
2	3342.5	96.8	2144	2 Q9S4M8	Q9S4M8 streptococc
3	453	13.1	1946	2 Q48545	Q48545 lactobacill
4	438.5	12.7	1902	2 Q9A1O2	Q9A1O2 lactococcu
5	424.5	12.3	1849	2 Q9SAK2	Q9SAK2 lactobacill
6	410.5	11.9	1585	2 Q9F8Q4	Q9F8Q4 streptococc
7	361	10.5	1647	2 Q9A180	Q9A180 streptococc
8	340	9.8	692	2 Q9EXK0	Q9EXK0 bacillus ps
9	298.5	8.6	757	2 Q9K6G6	Q9K6G6 bacillus ha
10	286.5	8.6	1150	2 Q53637	Q53637 streptococc
11	285.5	8.3	799	2 Q9KEM1	Q9KEM1 bacillus ha
12	279.5	8.1	1134	2 Q9L508	Q9L508 streptococc
13	257.5	7.5	1130	2 Q9K2K8	Q9K2K8 streptococc
14	256	7.4	731	2 Q9S547	Q9S547 cucumis mel
15	252.5	7.3	840	10 Q9F112	Q9F112 arbidopsi
16	251.5	7.3	1433	2 Q45616	Q45616 bacillus su
17	250	7.2	801	10 P93321	P93321 lycopersico
18	246	7.1	746	10 Q9LMA3	Q9LMA3 lycopersico
19	242	7.0	754	10 Q65834	Q65834 lycopersico

20	241	7.0	809	10 Q9SPA0	Q9SPA0 lycopersico
21	239.5	6.9	747	10 Q82007	Q82007 lycopersico
22	238.5	6.9	745	10 Q96478	Q96478 lycopersico
23	238	6.9	813	10 Q43546	Q43546 lilium long
24	234	6.8	736	10 Q9FYC7	Q9FYC7 oryza sativ
25	232.5	6.7	743	10 Q9LMA4	Q9LMA4 lycopersico
26	232	6.7	856	10 Q9SUN6	Q9SUN6 arbidopsi
27	231.5	6.7	816	10 Q9S4V5	Q9S4V5 arbidopsi
28	229	6.6	713	10 Q9F1M5	Q9F1M5 arbidopsi
29	228	6.6	747	10 Q65836	Q65836 lycopersico
30	227	6.6	754	10 Q65835	Q65835 lycopersico
31	223	6.5	747	10 Q9S4N2	Q9S4N2 lycopersico
32	220.5	6.4	666	10 Q9Z4R6	Q9Z4R6 lycopersico
33	219.5	6.4	766	10 Q9LZS6	Q9LZS6 arbidopsi
34	219.5	6.4	815	10 Q64481	Q64481 arbidopsi
35	212.5	6.2	749	10 Q9L1L8	Q9L1L8 arbidopsi
36	211.5	6.1	761	10 Q82777	Q82777 lycopersico
37	208	6.0	783	10 Q04190	Q04190 arbidopsi
38	206.5	6.0	697	10 Q9F1G1	Q9F1G1 arbidopsi
39	206	6.0	779	10 Q9LNU0	Q9LNU0 arbidopsi
40	204.5	5.9	741	10 Q9FCU3	Q9FCU3 arbidopsi
41	203.5	5.9	777	10 Q9ZS44	Q9ZS44 lycopersico
42	202.5	5.9	693	10 Q9F1M8	Q9F1M8 arbidopsi
43	202	5.8	745	10 Q04678	Q04678 lycopersico
44	198.5	5.7	707	10 Q9FEU4	Q9FEU4 arbidopsi
45	196	5.7	706	10 Q81324	Q81324 arbidopsi

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	2119 AA.
ID	Q9AHT5			
AC	Q9AHT5:			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	SERINE PROTEASE (FRAGMENT).			
GN	PTA.			
OS	Streptococcus pneumoniae.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=1313;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-M;			
RX	MEDLINE=2116976; Pubmed=11179332;			
RA	Wizemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,			
RA	Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,			
RA	Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,			
RA	Langermann S., Johnson S., Koenig S.;			
RT	"Use of a whole Genome Approach To Identify Vaccine Molecules			
RT	Affording Protection Against Streptococcus pneumoniae Infection."			
RL	Infect. Immun. 69:1593-1598(2001).			
DR	EMBL; AF291699; AAK19159.1; -			
KW	Protease.			
FT	NON TR.			
SO	SEQUENCE	1	1	
		2119 AA;	238226 MW;	517F9BF6B960A6A CRC64;

Query Match 100.0%; Score 3454; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 5.2e-174;
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1 YPVVLADTSSSEDALNISDEKVAENKEHENHISAMETSDQFEKKTAIVIKKEVSKN 60	
DB	1 YPVVLADTSSSEDALNISDEKVAENKEHENHISAMETSDQFEKKTAIVIKKEVSKN 60	
QY	61 PYVDNNTSNEBAKTKENSNSOGDYTDSPFNKNTENPKEDKVVYIAEFKDSGEKAI 120	
DB	61 PYVDNNTSNEBAKTKENSNSOGDYTDSPFNKNTENPKEDKVVYIAEFKDSGEKAI 120	

```

QY 121 KELSLSKNTKLYTYDRJFNGSALETTTPDNLKIKQIEGSISSVERAQKQVPMNHARKEL 180
DB 121 KELSLSKNTKLYTYDRJFNGSALETTTPDNLKIKQIEGSISSVERAQKQVPMNHARKEL 180
QY 181 GVEEAIDYLSKINAPFGKFKEDGRGVISNIDGTQYRHKAMRIDDDAKASMRKKEDLKG 240
DB 181 GVEEAIDYLSKINAPFGKFKEDGRGVISNIDGTQYRHKAMRIDDDAKASMRKKEDLKG 240
QY 241 TDKNWLSDKIPHAENVYNGKITVEKYDDGRDYPDPHGMHAGIAGNDEDDIKNFNG 300
DB 241 TDKNWLSDKIPHAENVYNGKITVEKYDDGRDYPDPHGMHAGIAGNDEDDIKNFNG 300
QY 301 IDGIAPNAQIFSYKMYSDAGSGFAGDETFHAIEDSIKHNVDVSVSSGFTGTGLVGEKY 360
DB 301 IDGIAPNAQIFSYKMYSDAGSGFAGDETFHAIEDSIKHNVDVSVSSGFTGTGLVGEKY 360
QY 361 WQAITALKRAGIPVYVATGNVATSSSSWDLVANNHLMKMTDGTGNTRTAAHEDALAVAS 420
DB 361 WQAITALKRAGIPVYVATGNVATSSSSWDLVANNHLMKMTDGTGNTRTAAHEDALAVAS 420
QY 421 AKNOTVEPKVNIAGESPFRNIGAFPOKSKITTTNEDGTAKPSKLEFVYIGKQDDDLG 480
DB 421 AKNOTVEPKVNIAGESPFRNIGAFPOKSKITTTNEDGTAKPSKLEFVYIGKQDDDLG 480
QY 481 LDLRKLIAMDRITTKDLKNAEKKAMDKGARAIMVYVTVNYNRDMWTELPAMGYEADG 540
DB 481 LDLRKLIAMDRITTKDLKNAEKKAMDKGARAIMVYVTVNYNRDMWTELPAMGYEADG 540
QY 541 TKSQVFSISGDDGVKLMNINPDKTEYVRNNKEDFKDLBOYYPIDMESFNKNRNGVD 600
DB 541 TKSQVFSISGDDGVKLMNINPDKTEYVRNNKEDFKDLBOYYPIDMESFNKNRNGVD 600
QY 601 EKEIDFKAPDPTDKELKEDITVPAGSTSMGPRIDLLKPDVSAFKNKIKSTLVNINGS 660
DB 601 EKEIDFKAPDPTDKELKEDITVPAGSTSMGPRIDLLKPDVSAFKNKIKSTLVNINGS 660
QY 661 TYGY 664
DB 661 TYGY 664

```

```

RESULT 2
Q9S4M8 PRELIMINARY: PRT: 2144 AA.
ID Q9S4M8
AC Q9S4M8:
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CELL WALL-ASSOCIATED SERINE PROTEINASE PRECURSOR PRTA.
GN PRTA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3.B;
RA Beche G., ten Thoren E., Bongaeerts R.J.M., Heinz H.-P., Zysk G.;
RT "Cloning and sequencing of a novel surface protease of Streptococcus
RT pneumoniae."
RL Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF17143; AAD48399.1; -.
DR HSSP: P07518; IMDE.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR003137; PA.
DR InterPro: IPR000209; Peptidase_S8.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF00882; Peptidase_S8; 2.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.

```

```

DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
KM SIGNAL.
FT SIGNAL. 1 19 POTENTIAL.
FT CHAIN 20 2144 CELL_WALL-ASSOCIATED SEI
FT PRTA.
SQ SEQUENCE 2144 AA: 240725 MW; 2052511470741331 CR...

Query Match 96.88; Score 3342.5; DB 2: Length 2144;
Best Local Similarity 96.38; Pred. No. 4.1e-168;
Matches 642; Conservative 14; Mismatches 6; Indels 5; Gaps 1;

QY 3 VVLADTSSSEDALISDREKY-----AENKEHENHSAMETSDFKKKTAVIKEKEV 57
DB 24 VVLADTSSSEDALISDREKYVVDKEKKNKHNDHAIETSDTEKKTIIIEKEV 83
QY 58 SKNPVIDNNTSNEBAKKEENSNSQGDYDSEVYNKNTENPKKEDKVVYIAEFKDSGE 117
DB 84 SKNPVIDTKTSNEBAKKEENSNSQGDYDSEVYNKNTENPKKEDKVVYIAEFKDSGE 143
QY 118 KATKELSLKNTKLYTYDRJFNGSALETTTPDNLKIKQIEGSISSVERAQKQVPMNHAR 177
DB 144 KATKELSLKNTKLYTYDRJFNGSALETTTPDNLKIKQIEGSISSVERAQKQVPMNHAR 203
QY 178 KEIGVEEAIDYLSKINAPFGKFKEDGRGVISNIDGTQYRHKAMRIDDDAKASMRKKED 237
DB 204 KEIGVEEAIDYLSKINAPFGKFKEDGRGVISNIDGTQYRHKAMRIDDDAKASMRKKED 263
QY 238 LKGTDKRWYLSDKIPHAENVYNGKITVEKYDDGRDYPDPHGMHAGIAGNDEDDIKN 297
DB 264 LKGTDKRWYLSDKIPHAENVYNGKITVEKYDDGRDYPDPHGMHAGIAGNDEDDIKN 323
QY 298 FNGIDGIAPNAQIFSYKMYSDAGSGFAGDETFHAIEDSIKHNVDVSVSSGFTGTGLV 357
DB 324 FNGIDGIAPNAQIFSYKMYSDAGSGFAGDETFHAIEDSIKHNVDVSVSSGFTGTGLV 383
QY 358 EKYWQAITALKRAGIPVYVATGNVATSSSSWDLVANNHLMKMTDGTGNTRTAAHEDAL 417
DB 384 EKYWQAITALKRAGIPVYVATGNVATSSSSWDLVANNHLMKMTDGTGNTRTAAHEDAL 443
QY 418 VASAKNOTVEPKVNIAGESPFRNIGAFPOKSKITTTNEDGTAKPSKLEFVYIGQDD 477
DB 444 VASAKNOTVEPKVNIAGESPFRNIGAFPOKSKITTTNEDGTAKPSKLEFVYIGQDD 503
QY 478 LIGDLRGKIAVMDRITTKDLKNAEKKAMDKGARAIMVYVTVNYNRDMWTELPAMGYEA 537
DB 504 LIGDLRGKIAVMDRITTKDLKNAEKKAMDKGARAIMVYVTVNYNRDMWTELPAMGYEA 563
QY 538 DEGTKSQVFSISGDDGVKLMNINPDKTEYVRNNKEDFKDLBOYYPIDMESFNKNRPN 597
DB 564 DEGTKSQVFSISGDDGVKLMNINPDKTEYVRNNKEDFKDLBOYYPIDMESFNKNRPN 623
QY 598 VGDEKETDFKAPDPTDKELKEDITVPAGSTSMGPRIDLLKPDVSAFKNKIKSTLVNIN 657
DB 624 VGDEKETDFKAPDPTDKELKEDITVPAGSTSMGPRIDLLKPDVSAFKNKIKSTLVNIN 683
QY 658 GKSTYGY 664
DB 684 GKSTYGY 690

```

```

RESULT 3
Q48545 PRELIMINARY: PRT: 1946 AA.
ID Q48545
AC Q48545:
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PROTEINASE (PRTB) PRECURSOR (PRTB).
GN PRTB.
OS Lactobacillus delbrueckii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;

```


OC Lactobacillus.
 OX NCBI_TaxID=1584;
 RN [1]
 RX SEQUENCE FROM N.A.
 RA MEDLINE=6236017; PubMed=8655480;
 RA Gilbert C., Atlan D., Blanc B., Portalier R., Germond J.E.,
 RA Lapierre L., Mollet B.;
 RT "A new cell surface proteinase: sequencing and analysis of the *prtB*
 RT gene from *Lactobacillus delbrueckii* subsp. *bulgaricus*.";
 RL J. Bacteriol. 178:3059-3065(1996).
 DR EMBL: LA8487; AAC41529.1;
 DR HSSP: P00782; 1SU6.
 DR InterPro: IPR003137; PA.
 DR InterPro: IPR00209; Peptidase_S8.
 DR Pfam: PF00225; PA; 1.
 DR Pfam: PF00082; Peptidase_S8; 3.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASIP; UNKNOWN_1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 DR SIGNAL.
 DR CHAIN 1 34 POTENTIAL.
 DR SIGNAL 193 1946 PROTEINASE.
 DR SEQUENCE 1946 AA; 212315 MW; 21EF1D02E79C6A0 CRC64;

Query Match 13.1%; Score 453; DB 2; Length 1946;

Best Local Similarity 25.7%; Pred. No. 1.6e-15;
 Matches 175; Conservative 114; Mismatches 261; Indels 130; Gaps 28;

OY 5 LADTSSSEDALNIDSEKVENENKHEHNIHSAETSDQFEKKT-----AVIKEEYVSK 59
 DB 27 LAGTFOGAFQOASQETSPRSASRAALTKYLOEQRYNNAKSKQFEAKEDQOASG 86
 OY 60 NPVIDNNTSNEAKIKENSKSG-DYTSFVNKNTEENPKKEDKYVYIAEFKESGEK 118
 DB 87 QAVSKKNESSRVIV---SLKSAFPHIT---SKPTGSAASVAKIKESQASD-QVAKDGEK 138
 OY 119 AIKELSLKNTKVLTYTDFNGSAIETTPDNLKIKOIEGSSVERAKQVPMNNHARK 178
 DB 139 VIKOVEEITGNKVRQGLVNAFSDIMDDIDKVDLPQVKNVTPRYKVP----- 191
 OY 179 EIGVEEADYLSINAPR-GNEDGRGVTSNIDTGDYRHKAMRIDDD-----AKASMR 232
 DB 192 ---DDEASQMAQVODWQEQEKLKGEGVTSIIDTGIDSSHQDLKDSGVSTALSKSEVE 248
 OY 233 FKREDLGTGKNWYLSKIPRAFNYYNGKRTVEKYDDGRYRPHGHNIAGLANTGE 292
 DB 249 SDKSKL-GHGKYY--TEKVPYGVYADKNDQIV---DNCGG--EMHGQHVAGIAGAN-- 297
 OY 293 QDIKNFNGIDGIAIPNAOIFYKMTS-DAGSGFAGDETFMHAIEDSIKHNVDVSSGFT 351
 DB 298 -----GVKGVAPDADQDLAMKVFSSNNAKNGAYDDOLISIEDSVKIGAVINMSLSGV 351
 OY 352 GTGL-VEKTYWQAIKALRKAGIPVAVATGATYATSSSSWDLVAN--NHLKMTDTGNT 407
 DB 352 SSDDGPDPQOQAAKASAEVINVISAGNGVAGSTADGNPNVNTGSELSTVGTPGV 411
 OY 408 RTAHEDAIIVASAKNOVREDKY--NIGSESEKRN--IAAFDPSKITTNEGCTKA-- 461
 DB 412 -----PALTVASAENSKVTTDTVKDELGVTFSSNSLKA---AQVTTQLSNNSVL 462
 OY 462 PSKLKPYVIGGQDQDLIG--LDLRGIAMDR-IYTKDLNAFKKAMDGARAIVNVN 517
 DB 463 TKKTLKLVDMGAGADYTAEKKAKEVGQGLAVYKRGATYFSAKAVANAKA--AGAAGIYIYN 520
 OY 518 TVNYNNDNTTELPAMGEADEGRTKSOVSTISGDDGVKLMNMINPDKKTEVRKNNKDEK 577
 DB 521 -----SEDDG--LLSMSLDDKTFPTLGSKADGK 547
 OY 578 DKLEOYYPIDMESFNNSKNPNYGD---KEIDFKAPDTRDKLYKEDIIVPAGSTSGMPRI 634
 DB 548 FWLQOQKKVRAKRLKFTALIDNSRAGKMSDF-----TSMGPTP 586

OY 635 DLLKPDVSAAPKNIKSTLN 654
 DB 587 ELDERPETAAGKXIYSLAN 606

RESULT 4

ID 09A102 PRELIMINARY; PRT; 1902 AA.

AC 09A102;

DT 01-JUN-2001 (Tremblrel, 17, Created)

DT 01-JUN-2001 (Tremblrel, 17, Last sequence update)

DT 01-JUN-2001 (Tremblrel, 17, Last annotation update)

DE PRT PRECURSOR.

GN PRT.

OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).

OC Plasmid PHP003.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OX NCBI_TaxID=1359;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=HP;

RA Christensen C., Piliidge C.J., Ward L.J.H., O'Toole P.W.;

RT "Characterization of a lactococcal plasmid in *Lactococcus lactis* subsp.

RL submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF247159; AAK27981.1;

KW Signal: Plasmid.

FT SIGNAL 1 33 POTENTIAL.

SC SEQUENCE 1902 AA; 200497 MW; 50B3D9DAE575B2B CRC64;

Query Match 12.7%; Score 438.5; DB 2; Length 1902;

Best Local Similarity 27.3%; Pred. No. 9.1e-15;
 Matches 166; Conservative 91; Mismatches 239; Indels 111; Gaps 25;

OY 78 NSNRSGDGYTSFVNKNTEENPKED--KYVYIAEFKDESEKAIKELSLKNTKVLTY 134
 DB 84 NKVQOQDIYVYIOMSA--PASENGTLRTDYSSTAEIQETNKIYAQASVKAVEGVT 142
 OY 135 -----YDRIFNSAITTPDNLKIKOIEGSSVERAKQVPMNNHARKIEIGVEEAD 187
 DB 143 QQTAGESGVYVNFSTVRVVDIPKLOAGVKTVTLAKYYPDADAKANSMAVQAVWS 202
 OY 188 YLSINAPFGKNGRGVTSNIDTGDYRHKAMRIDDAKASMKKEDL-KGTD--K 243
 DB 203 NYK-----YKGEGVTSYIDSGIDPTHKDKRLSD--KDKVLTSDVEKFTDYK 251
 OY 244 NYWLSDKIPRAFNYY-NGKRTVEKYDDGRYRPHGHNIAGLANTGEDIKNFNGID 302
 DB 252 GRYENSKVPYGFNVADNNDITDKVD-----EGHGHVAGIIGANGTGD--PAKSV 303
 OY 303 GIAPNAQIFSTKMTSDAG-SCFAGDETFMHAIEDSIKHNVDVSSGFT-GTGLVGEKY 360
 DB 304 GVAPEAQDLAMKVFSTNSDTSTGSDTLVSAIEDSAKIGAVULNMSLSGSGNQLDEPE 363
 OY 361 WQAIKALRKAGIPVAVATGATYATSSSSWDLVANNNHLMKMTDTGNTVTAHEDAIIVAS 420
 DB 364 IAAVONANESGTAIVISAGNSGTSAT--EGVKNQDYGIDDNENGVGPGTSKGAITV 421
 OY 421 AKNOVREDKYNI-----GGESEFKY--RNIGAFDPSKITTNEGCTKAPSKLKYI 470
 DB 422 AENDVITQAVTITDGTGLGPEITQLSSNDPFGSFQKKFYVYKDAAGNLSK----- 475
 OY 471 GKQDQDLIGDLNGKIAVMDR-IYTKDLNAFKKAMDGARAIVNVTVNYNNDNTTE 529
 DB 476 GKVADYT---ADAGKTAIVKRGELTFEDDKQKYQA--AGAAGLIYN-----NDGTATP 525
 OY 530 LPAMGEADEGRTKSOVSTISGDDGVKYL--NMNINPDKKTEVRKNNKEDFKDLQGYPID 587
 DB 526 VTSMALT---TTPPTFELSSVTGKLVQVYTAHPDSDLSGK----- 563
 OY 588 MESFNNSKNPNYGDKEIDFKAPDTRDKLYKEDIIVPAGSTSGMPRIIDLKPDVSAAPK 647

```

Db 564 -----IALLTVPN---QKYTEDKM--SDFTSYGVPVNSLFRPDITAPG 602
Oy 648 NIKSTLN 654
Db 603 NIMSTON 609

RESULT 5
O9S4K2 PRELIMINARY: PRT: 1849 AA.
ID O9S4K2.
AC O9S4K2.
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CELL ENVELOPE-ASSOCIATED PROTEINASE.
GN PRTH.
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OC NCBI_TaxID=1587;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CNR232;
RX MEDLINE=99350424; PubMed=10419958;
RA Pederson J.A., Milleki G.J., Welmer B.C., Steele J.L.;
RT "Genetic characterization of a cell envelope-associated proteinase
RT from Lactobacillus helveticus CNR232."
RL J. Bacteriol. 181:4592-4597(1999).
DR EMBL: AF133727; AAD50643.1; -.
DR HSSP: P00782; 150C.
DR InterPro: IPR001444; Flag_db_rnd.
DR InterPro: IPR003137; PA.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 3.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 1849 AA; 199614 MW; 219F0D44B15A091F CRC64;

Query Match 12.3%; Score 424.5; DB 2; Length 1849;
Best Local Similarity 24.1%; Pred. No. 4,8e-14;
Matches 160; Conservative 113; Mismatches 258; Indels 133; Gaps 25;

Oy 31 ENHSAMETSDPFKEKKAIVAIKEKEVNSKNPYIDNNTSNEEAKIKEENSNSKSGDITDSF 90
Db 27 QOVKASVDSQRTYVEKSTKAEESTANLTKAVEQLAKGVNFHLTVNQKQDQVYVDVI 86
Oy 91 V-----NKNTENPKEDKVVYIAEFKDESGEKAIKE-ISSLKMTKVLV 133
Db 87 VOLSTPRAATNGSVANSSSAEIEDASKKVI-----ANQASIKKVKAITMOALGK 137

Oy 134 TYDRIFNGSAIETTPDNLKIKQIEGISSVERAOKVOPMMNHARKEIGVEEAIDVLKSTN 193
Db 138 SYGVYVNGFATRAKAKVKDQIKLRNIPGVKSVTLAK-----VYVAN-----DSDADMANVNS 187

Oy 194 APFGK-NPDGGMWYSNIDGTGYRHKAMRIDDDAKASMRFEKEDLKGTDKRY--WLSDK 250
Db 188 TVMNNYKKKGEGTVVSIIDGTIDPNHMKDLRLSDSKVKLTQKVNAAFTGESYGRYFTDK 247

Oy 251 IPHAFNXY-NGGKITVEKVDGROVDFPHGMHAGIILAGNPDIEDIKNPNNGIDGIAPNAQ 309
Db 248 VPTGHNSDNDNNT---DNDPS--EAGHGMHAGIIVANGTADSV---NSVGAAPAQ 298

Oy 310 IFSYKMYSDAGSGFAGDET-MFHAIEDSIKHNVDVAVSSG-FTGGLGVEGYMQAIRAL 367
Db 299 LLAKKAFENSDDASSTDSITIGALDSDAKGADVLLNMSLSGVSEQEDDEVAVAVERA 358

Oy 368 RKACIPMVVATGNGVATSASSSSWDLVANNHLMKMTDGTGVTFTAHHEDAIAVASAKNGTVE 427

```

```

Db 359 TRKGTAIVISAGNSGT--SNSIEGVKAYYGNDPMETLGNPGTARSATTVAASAKTKAT 416
Oy 428 EDKYNIGGESFKNRNG-----AFPKSKITTEDGTAKPSKLPYIGGOD 475
Db 417 TDGVTTTSADGKTTIAGPEATQLEGTDRAPFNDKRYVVKD-----KNGNLGTGSA 468
Oy 476 QDLIGLDRGKIVMDR---IYTKDLNAPKAKMDKAGARAIWVNYVYRNDRMTLPA 532
Db 469 KQYTSA-VKCKIIVKRGELTF-----DKQKVAQENAGACLIIVN-----NKAGDITG 516
Oy 533 MGEYADEGTSQVFSISGDDGVKLWNI--NPKKTEVRKRNKEDFKLEQYYPIDMES 590
Db 517 MLNNAEFPFA---GLSATSGEKLVKVEAHPEALKVS-----IYVQA 556
Oy 591 FNSKRNNGDEKIDFKFADPTDKELYKEDIYPAGSTSGPRDILLKPDVASPKGNK 650
Db 557 LNNS-----AKQTD-----LMSDFTSYGVPVNSLAFKPDISAPGSHW 593

Oy 651 STLN 654
Db 594 STON 597

RESULT 6
O9F8Q4 PRELIMINARY: PRT: 1585 AA.
ID O9F8Q4.
AC O9F8Q4.
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CELL ENVELOPE PROTEINASE.
GN PRTH.
OS Streptococcus thermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1308;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20509797; PubMed=11055922;
RA Fernandez-Espia M.D., Garault P., Monnet V., Rul F.;
RT "Streptococcus thermophilus Cell Wall-Anchored Proteinase: Release,
RT Purification, and Biochemical and Genetic Characterization."
RT Appl. Environ. Microbiol. 66:4772-4778(2000).
DR EMBL: AF243528; AAG09771.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR003137; PA.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 3.
DR PRINTS: PR00723; SUBTILISIN.
KW Envelope protein.
SQ SEQUENCE 1585 AA; 169873 MW; AD210BCE1F9A399 CRC64;

Query Match 11.9%; Score 410.5; DB 2; Length 1585;
Best Local Similarity 24.3%; Pred. No. 2,1e-13;
Matches 160; Conservative 104; Mismatches 245; Indels 149; Gaps 24;

Oy 44 KEKKAIVAIKEKEVNSKNPYI-----DNTSNEEAKIKEENSNSKSGDITDS 89
Db 2 KKEKFSLRKKYKIGTVSVLLGAVFLPAGAPVADELTSIVETKRV-----EATVPDA 53

Oy 90 FVNKN-TEPKKEDKVVYIAE-----FKDKSGEKAIKELSLKNTKVLTYTDRFNG 141
Db 54 IYSESASESPVVEELVDTSVEATSTVYTTDNEETPGSEALENSANTVEETQ----- 107

Oy 142 SAIEETPDNLKIKQIEGISSVERAOKVOPMMNHARKEIGVEEAIDVLKSTNAP--FGKN 109
Db 108 PAVETPAISEKVEEKEKISVADETTAI-----TNOEAKPPONIDSNITITTPKMYSG 161

Oy 200 PDGRGVNISNIDGTGYRHKAMRIDDDAKASMRFEK-----DLKGTDKNYWLSDKIPHA 254

```

QY	Db	162	YKGGSTVAIIIDSGLDVHDVLIHISDLSSTAYKYSKSELEAKGAGIYYGEFNDKXVFG	221
QY	225	FNYYNGGKIYIEKIDGDGADYDFDPHGMHITAGIACDTEODIKNFNGIDGIAPNAOIFSTK	314	
Db	222	YNYVDVN--TVLKEDKRS---HGMAHTSIATGNPTOPVAGQL--MYGVAPEAOVMFMR	273	
QY	315	MYSDAGSGFADDETFMFIHIEDSIKHNVDVVSAGSGSTGTGLV--GEKXWQAIARLRKAGI	372	
Db	274	VFSIDL-KATTGAALYKKAIEDVYKIGADUSINISLGGANGSVYNNMENVTALEAARRAGV	332	
QY	373	PMVATGCVATYASSSSMDLVANMLKMTDIGNVYRTAAHEDAIYVASAKNOTVEFDVYN	432	
Db	333	SVYIAGAGDGFSGSHS-----NP SADYPDYGLVAPSTAHDAISVASYNNNTVSKEYIN	387	
QY	433	IGGSESKYRNIGAFEDSKITTTNEDGTAP-----SKIKFYIGKGGDDQDGLDLNGK	486	
Db	388	IIGLE--NNADLNYGKSSFPDPE--KSPVFEIGKEYEYVYAGISGADPDGDLTGK	441	
QY	487	IAYMNR---ITTKDLKNFKKAMDKGARAIVNVN-----VNYNRRDNTWTLPMGYEA	537	
Db	442	LALIKGTITTSSEKIAN---ATAAGAAGVYVLENSPCGBANYSKMDLDPAIYAPSV---	493	
QY	538	DEGTSQVFSISGDDGVYKLMNMINPDKTEVYKRNKEDFKDLQYYPIDM-ESFNSNKP	596	
Db	494	-----FILEFEBALAANSY	508	
QY	597	NYGDEKEIDFPKAPDTOKELYKEDIIVPAGSTWGPRIIDLKLPVSAAGKNIKSTLN	654	
Db	509	KIAFNENETDIR--PAPKAGLLSD-----FSSWGLSDAGELKPLDAAPGAIYVAIN	557	
RESULT	7			
Q9A180	ID	PRELIMINARY;	PRT, 1647 AA.	
Q9A180:				
AC	09A180:			
DT	01-JUN-2001 (TREMBLrel, 17, Created)			
DT	01-JUN-2001 (TREMBLrel, 17, Last annotation update)			
DT	01-JUN-2001 (TREMBLrel, 17, Last annotation update)			
DE	PURATIVE CELL ENVELOPE PROTEINASE.			
OS	Streptococcus pyogenes.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
CC	Streptococcus			
OX	NCBI_TaxID=1314;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN-SF370;			
SC	MEDLINE=21192684; PubMed=11296296;			
RA	Ferretti J.J., Meshin W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,			
RA	Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,			
RA	Yuan Y., Cila H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,			
RA	Qian X., Clifton S.W., Roe B.A., McLaughlin R.;			
RT	"Complete genome sequence of an M1 strain of Streptococcus pyogenes."			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).			
DR	EMBL; AE006503; AKR3444.1;			
SO	Envelope protein; Complete proteome.			
CO	SEQUENCE 1647 AA; 181287 MW; F36E6CB965C291A2 CRC64;			

	Query Match	Similarity	10.5%	Score 361	DB 2	Length 1647
	Best Local	Similarity	21.6%	Pred. No. 9,1e-11		
	Matches 156	Conservative	99	Mismatches 247	Indels 220	Gaps 24
QY	37	METSDPFKEKK-----		TAIVKEKENVSKNPVIDNNNSNEAKI--KE	76	
DB	1	MEKKRFRFSIRAKKSGTFSVLISGVFLVMTTYYAADELSTMSPTLTNNAAQQAHLITNE			60	
QY	77	ENSNSQSGDYDTSFVNR--NTENPKRKEDKVYIAEFKDKESGKAKELSLANTKVLTYTY			135	
DB	61	LSASAKSKSDTSQLILTKLRNKEKGQSDPL-----SEPTTELADPDASMANNT-----			108	
QY	136	DRIFNGSALETTTDDNLIDKIKOLEGJSSVERAOKVOPMMNHAKREIGVEAIDYLSINAP			195	

[illegible]

RESULT	8		
ID	Q9EXK0	PRELIMINARY;	PRT; 692 AA.
AC	Q9EXK0;		
DT	01-MAR-2001 (TREMBLrel. 16, Created)		
DT	01-MAR-2001 (TREMBLrel. 16, last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)		
DE	VPR [EC 3.4.21.] (FRAGMENT).		
OS	Bacillus pseudofirmus.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OX	Bacillus/Staphylococcus group; Bacillus.		
OX	NCBI_TaxID=79885;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-OF4;		
RX	MEDLINE=92017665; PubMed=1833620;		
RA	Ivey D.M., Krulwich T.A.;		
RT	"Organization and nucleotide sequence of the <i>atp</i> genes encoding the		
RT	<i>ATP synthase</i> from alkaliphilic <i>Bacillus firmus</i> OF4."		
RL	Mol. Gen. Genet. 229:292-300(1991).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-OF4;		
RA	Hicks D.B., Krulwich T.A.;		
RT	"Bacillus pseudofirmus OF4 <i>ATP synthase</i> (F1F0 <i>ATPase</i>) and upstream		
RT	region."		
RL	Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.		
DR	EMBL; AF330160; M6G4835.1; -. InterPro: IPR003137; PA.		

DR InterPro: IPR000209; Peptidase_s8.
 DR Pfam: PF02225; PA: 1.
 DR Pfam: PF00082; Peptidase_s8; 2.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
 KM Hydrolase.
 FT NON_TER
 SEQUENCE 692 AA; 75155 MW; E53DCB38E6B6D9 CRC64;

Query Match 9.8%; Score 340; DB 2; Length 692;
 Best Local Similarity 24.9%; Pred. No. 3.3e-10;

Matches 144; Conservative 77; Mismatches 150; Indels 208; Gaps 26;

QY 126 LKNTVLYTRIFNGSAIETTPDLKIKQIEGSSVERAKVQPMNHAKKEIGVEEA 185
 DB 1 LPSSKVRKKFNLYNGFSLSEKGLDVLKDYQ---SVGRVQV-----AQYEANLDGS 51
 QY 186 I-----DYKSTINADPGKNFGDGRGVNISNIDTGTDRHKAIRIDDAKASMRKKEDLKG 240
 DB 52 VPIGSDHVRGMFSRDEGLTGKGVAVIDTIGIDYRH-----PDLQA 94
 QY 241 TDKNWLSDKIPHANVYNGCKITVEKYDDGDRYDP-----HGMHAGILAG 288
 DB 95 N-----KKGGDYVD-YD---HDMETQSTQGVPTLHGTHVAGITAA 132
 QY 289 NDEODIKNFNGIDGIAFNAQIFSKYKMSDASGFGADETMFHAIEDSIKHNVDVSVSS 348
 DB 133 N-----GVKGVAPAEADIAIRALGPGQGITT--EQVIEALEKAVADVDVLANSL 181
 QY 349 GFTGTGLVGEKYMQAIRALKRAGITPMVAVATGNVATSSASSWDLVANHLKMTDTGNVTR 408
 DB 182 GNTVNG---PDMPSTVALDK----- 198
 QY 409 TAAHDAIAVASAKQGYEFQKVNIGSGSEFKYRNIG-----FEKSKITINE 456
 DB 199 -AVEGCVAVVTSNGNSGPMMTVSGPSTKAISVGASAPRIKTPYATVFGDKELTLP 257
 QY 457 DGTAPSKLK---FVYIYGKGDODLIGDLRGKIAVMDR---IYTKDLKNAFKKAMDG 509
 DB 258 MCGTAPMALKRDFPLVDGGLGTEEDLDELEVGSVLVLRGILPTPEKAHNA--KKA---G 313
 QY 510 ARAIVAVTVNTVYKNDNMTLP-AMGYEADGTSQVFSISGDDGVKLMNINPKKTEV 568
 DB 314 AKAMIIYN-----NLPGAFVGYEGVKLPVVSVTKEGD----- 347
 QY 569 KRNNEDEKDKLEQYYPIOMESFNKPNVDEKEIDFKFADDTKELK--EDIIIVAG 626
 DB 348 -----EYLDKLEAH-----KEL-----TIRTIYREDEDTIAPES 378
 QY 627 STSMGPRIDL-LKPDVAPGKNIKSTLN---VINGKS 660
 DB 379 SR--GPVQTWEVKPDLVAPGVSIDSTYPRGYALANGTS 415

RESULT 9
 Q9K6G6 PRELIMINARY; PRT; 757 AA.
 AC Q9K6G6;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE MINOR EXTRACELLULAR SERINE PROTEASE.
 GN BH3763.
 OS Bacillus halodurans.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 NCBI_Taxid=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C-125 / JCM 9153;
 RX MEDLINE-20512582; PubMed-11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis."
 RT Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL: AP001519; BAB07482.1; .
 DR InterPro: IPR001337; PA.
 DR InterPro: IPR000209; Peptidase_s8.
 DR Pfam: PF02225; PA: 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KM Protease; Complete proteome.
 SEQUENCE 757 AA; 82516 MW; F6E28CF597435E1 CRC64;

Query Match 8.6%; Score 298.5; DB 2; Length 757;
 Best Local Similarity 21.6%; Pred. No. 5.9e-08;

Matches 127; Conservative 97; Mismatches 181; Indels 183; Gaps 25;

QY 96 ENPK---KEDKVVYIAEFKDKESGKAIKELSSKNTKLYLYDRIFNLSALETTPDND 152
 DB 30 EKPRDLADEYVIVLTJNDLETITIKDYER--SISGELRMTE--RLNGFSFQLEPEEVE 86
 QY 153 KIKQIEGSSVERAKVQPMNHAKKEIGVEEAIDYLSINAPFGKNFGDGRGVNISNIDT 212
 DB 87 KLTYSVGRVDHVMVTYEATINESAPFIGAQD---VRQMDDEGVHLTGKVKKAVIDT 142
 QY 213 GTDYHKKAMRIDDAKASMRKKEDLKTGNWYLSDKIPHANVYNGCKITVEKYDDGR 272
 DB 143 GIDYTH-----PDLQSSYKGGYDFVD-YDD-- 166
 QY 273 DYFDD-----HGMHAGILAGNDFQDILKNFGIDGIAFNAQIFSKYKMSDAG 320
 DB 167 ---DPMETIASGCPPTLHGTHVSGITAAH-----GVKGVAPAEADIAIRALGPG 214
 QY 321 SGFADDETMFHAIEDSIKHNVDVSVSGFTGTGLVGEKYMQAIRALKRA--GIPMYVA 377
 DB 215 QGTT--EQVIAIEKAVEGCVVINLSGNVYNG---PDWPTSLADAAVEGVAVATS 268
 QY 378 TGNVTSASSSSWDLVANHLKMTDTGNVTRFAHEDAIYV-ASKAKNOTVEEDKNIGGE 436
 DB 269 NGN---SGPNMW-----TVGSPGTSKKAISVGASAPPLNTPY----- 302
 QY 437 SPKYRNIGAFPKSKITTNEDGTRAP---SKLKFVYIYGKGDODLIGDLRGKIAVMDR 492
 DB 303 -----LTFGEENELISLPPSGGLPMAFKRDLPMIDVGYTEKEBEGVDAGKVVLIK 356
 QY 493 ---IYTKDLKNAFKKAMDGARAIVVTVNYNRDNMTLPAMGYEADGTSQVFSIS 549
 DB 357 GNVPTFEKVMH---AVAKAANGVLIYNN-----TPGFTGM-IEGCVNIIVYSIT 402
 QY 550 GDDGVKLMNMINPKKTEYK---RNKKEDEKDKLEQYYPIOMESFNKPNVDEKEID 605
 DB 403 REDGEFLLEOLELQKNEKELTIRTKEDF-----VALFSSGCP----- 442
 QY 606 FRFADPTKELKEDIIVPAGSTSMGPRIDLKLPDVAPGKNIKSTL 653
 DB 443 -----VHTWD-----VKPDVAVPGVSDSTI 464

RESULT 10
 O53637 PRELIMINARY; PRT; 1150 AA.
 AC O53637;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE SCPB.
 GN SCPB.
 OS Streptococcus agalactiae.

OC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-78-471;
RA Chmoungyulna I., Suvorov A., Cleary P.;
RL Infect. Immun. 64:0-0(0).
DR EMBL: U56908; AAB17762.1; -
DR HSSP: Q45670; IDBI.
DR MEROPS: S08_020; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR003137; PA.
DR InterPro: IPR00209; Peptidase_S8.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 4.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
SEQUENCE 1150 AA; 126309 MW; 7BA9366BEC71633C CRC64;

Query Match 8.6%; Score 296.5; DB 2; Length 1150;
Best Local Similarity 24.3%; Pred. No. 1.4e-07;
Matches 122; Conservative 78; Mismatches 163; Indels 139; Gaps 24;

OY 189 LKSNMPCGNPGRGVINIDGTGRHKAHRIODDASAKSRFKEDKGTDKN--- 245
DB 111 VKTLOEKAG---GAGTVAVVIDGFPKHEAMKLTDTKARYO-SKEDLEKAKKEGKIT 166
OY 246 ---WLSDKIPHAENYNGKLTVEKYDDGDRYDPHGMHAGILAGNDTODIKNFNGID 302
DB 167 YGEVNDKVAHYDYSDKGTAVDQ-----EHGTHVSGILSGN-ASEKREPERLE 216
OY 303 GIAPNAOIEFYK-----YSDAGSGFAGDETMHATDSIKHNVVSVSGFTGGLVG 357
DB 217 GAMPEDALLMREYVINGLADYARNA-----QATRDAINLGAKYIMSGMAALAYAN 270
OY 358 ---KRYMOAIALKRAKGIPIVAVATGNVATSSASSSMDLVANNHLKMTGVTTRAHEDA 415
DB 271 LPBETKAKAFYAKSKGVSIYTSAGNDSSFGCKTRPLA--DH--PDYGVGPAAADST 325
OY 416 IAVA---SAKQVTEFEDKVNIGSGEFKRNIGAFEDSKITTNEDGTAPSK-LKFVYIGK 472
DB 326 LTVASVSPKQLETVAVKTAADQDK-----EMPVLTNR--FEPKKAQVAYANR 374
OY 473 GQDODLIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVNVYVYNNRD----- 525
DB 375 GTEKEDF-KDYVKRIALIER-GDIDFCKLAKAKGAVGLI-----YNODKGFPIEL 427
OY 526 -NMTELPAMGYEADGCKSQVFSISGDDGVKLMNMINPDKTEKRRNKKDKPKLQYY 584
DB 428 PNVDMQMAA-----FISKQGLLKD--NP----- 450
OY 585 PIDMESFNKPNVGEKEIDFKFAPDTEKELYKEDIIVPAGS-----TSMGPRIIDL 637
DB 451 -----OKITFNATPK-----VLPTASTKLSRSSMGLTADGN 484
OY 638 LKPDVSAFGKNISTLVNIGK 659
DB 485 IKPDIAFGODILS--SVANNK 504

RESULT 11
O9KEM1 PRELIMINARY; PRT; 799 AA.
AC O9KEM1;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE MINOR EXTRACELLULAR SERINE PROTEASE (EC 3.4.21.).

GN VPR OR BH0831.
OS Bacillus halodurans.
OC Bacteria: Firmicutes; Bacillus/Clostridium group;
OX Bacillus/Staphylococcus group; Bacillus.
OC NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C-125 / JCM 9153;
RA MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.,
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001510; BAB04550.1; -
DR InterPro: IPR003137; PA.
DR InterPro: IPR00209; Peptidase_S8.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 2.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Protease; Hydrolase; Complete proteome.
SEQUENCE 799 AA; 86192 MW; AF0FF17C2BD34D8A CRC64;

Query Match 8.3%; Score 285.5; DB 2; Length 799;
Best Local Similarity 21.4%; Pred. No. 3.1e-07;
Matches 140; Conservative 96; Mismatches 196; Indels 221; Gaps 30;

OY 63 IDNNTSNEAKIRE-ENS----KSQGDYDSEFNKNTENPKKEDKVVYIAEFKDESGE 117
DB 43 IDTSSALFVYIEIDPSTIEAKHG-----QKSKNELKQAQSVYEQID----- 90
OY 118 KAIKELSLKNTVLYYDRIENGSAIETTDPDLNKTQIEGISSVERACKVOPMNNHAR 177
DB 91 -----LVPSSTVTHEYDELFGFALFLPAHQIPSTLIGDVHAV--YPNIEVEVETDF 141
OY 178 KEIGVE-----EADVLKSNAPF-GKN-----FDRGMVSNIDGTGRHKAARID 224
DB 142 DEVAIEEDAYSPEMLD-----SAPFICANDAMEAGYIGESITVYAILDTGVYTH----- 190
OY 225 DDAKASRFEKEDLKGDKNYWLSDKIPHAENYNGKLTVEKYDDGDRY--FDP----- 277
DB 191 -----PDLVHAFQYKKWDF-IDNNDPQEPTRPPDPRGIER 225
OY 278 -HGMHAGILAGNTEODIDKPNFGIDGIAIPNAQIFSTKMTSDAGSGFAGDETMFHAEIS 336
DB 226 THGTHVAGTVAAAGL-----IKGVAPDANLAVRVLGPGRGSTAG--VIAGIERA 274
OY 337 IKHNVDVSVSGFTGGLGKRYWQAIRALKRAGIPMVAVATGNVATSSASSSMDLVANN 396
DB 275 VQGDADIMNLSLGT-----LNDPFAVST-ALDM-AMAE 308
OY 397 HLKMTDGNVTRTAHEDAIYASAKQVY-----EFDKVNIGSGEFKRNIGAFEDSK 451
DB 309 VAVATVSGN-----SGPNMTVGSPTSDAISVGTALPYKYY-----KAS 350
OY 452 ITTNEDGTAPK-----KLFVYIGKQDODLIGDLRGKIAVMDRI 493
DB 351 VFTS-DEIDYPSADIMGFPSEDELELDEGTVEYAFAGLCKPGDFEGVDGKALALVRG 409
OY 494 YTKDLKNAFKKADKARALMVNTVYVYRDMMTELPAWGTEAD-EGTSQVFSISGDD 552
DB 410 EIPPEVKA-----ENKKAAGAVGALTYN-----VAGVQYVPGLAIPITMLNED 455
OY 553 GVKLMNINPDKKTEYVRNKKEDFKDLQYYPIDMESFNKPNVGEKEIDFKFAPDT 612
DB 456 GLMKRNLENGQNT-----VTFSIEF 476
OY 613 DKELYKEDIIVPAGSTVSGPRIDL-LKPDVSAFGKNIKSTLVNIGKSTYGY 664

Db 477 DK-LVGETV---ADFSRRGPMHTWIKRPDVSAFVAIVSTRIPRQDPDPYGY 525

RESULT 12

Q9L908 PRELIMINARY; PRT; 1134 AA.
 AC Q9L908;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE C5A PEPTIDASE (FRAGMENT).
 GN SCPB.
 OS Streptococcus agalactiae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1311;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=130;
 RA Adderson E.E., Bohnack J.F., Takahashi S.;
 RT "Genetic Polymorphisms of Group B Streptococcus agalactiae scpb."
 RL Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF189004; AAF04282.2; -
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR00209; Peptidase_S8.
 DR Pfam; PF02225; PA.1.
 DR Pfam; PF00082; Peptidase_S8; 4.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 FT NON_TER 1134 1134
 SQ SEQUENCE 1134 AA; 124414 MW; 723B8DC8B330B73 CRC64;

Query Match 8.1%; Score 279.5; DB 2; Length 1134;
 Best Local Similarity 24.3%; Pred. No. 1.1e-06;
 Matches 122; Conservative 75; Mismatches 167; Indels 139; Gaps 25;

Qy 188 YLKSINAPFGNPDGRGVINIDTGYRHKAMRIDDDAKASMKKEDDKTKKNY-- 245
 Db 110 HVKTLQEKAGK--GAGTVAVVIDAGFKNHEAMRLDTKARYQ-SKENLEKAKKEHGI 165
 Qy 246 ---WLSDKIPAHFYNGKITEVEKYDDGRDYFDPGHMIGIAGNTEODIKNFNGI 301
 Db 166 TYGEWVNDKVAIYHDSKDGKNAVDQ-----EGHTVSGILSGN-ASEMKKEPYRL 215
 Qy 302 DGIAPNAQIFSYKM---YSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLV 356
 Db 216 EGAMPEAOQLLMRVEIVNGLADYARNYA-----QAIRDAVNLGAKVIMNSFGNALAYA 269
 Qy 357 G--EKYWAIRALKRAGIPMVVATGNVATSASSSSMDLVANNHLMKTGTGNTRTAAHED 414
 Db 270 NLPDETAKAFDYAKSGVSIYTSAGNDSFGKPRPLPLA--DH--PDYGVVGTTPAAADS 324
 Qy 415 AIAVA--SAKNQTFEEDKNIGCESFKRYRNIGAFEDSKITTNEDGTAPSK-LKFEVYIG 471
 Db 325 TLTVASYSPOQLTETATVKTDDHDK-----EMPYISTNR---PEPKAIYAYAN 373
 Qy 472 KGQDODLIGDLRGKIAVMDRIYTRDLKNAFKKAMDKGARAIMVNTVYVYNRD----- 525
 Db 374 RGTKEDDF-KDVEGKIALIER-GDIDFKDKIANAKKAGAVGLI-----YNDQKGFPIE 426
 Qy 526 ---NMTLEPAMGYEADGKTSQVFSISGDGVLKLMNMINPDKITEVKRNKKEDFKKLEQY 583
 Db 427 LPNDQMPAA-----FTSRDGLL-----KD----- 448
 Qy 584 YPIDMESFNKPNVGEKEIDFKFAPDDELYKEDIIVPAGS-----TSMGPRIDL 636
 Db 449 -----NPP-----KITTFNATPK-----VLPTASGTKLSRFSWGTLTADG 483
 Qy 637 LKPDVSAPEGKNIKSTLVINGK 659

Db 484 NIKPDIAAGQDILS--SVANNK 504

RESULT 13

Q9K2K8 PRELIMINARY; PRT; 1130 AA.
 AC Q9K2K8;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE C5A PEPTIDASE (FRAGMENT).
 GN SCPB.
 OS Streptococcus agalactiae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1311;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=125, AND GW;
 RA Adderson E.E., Bohnack J.F., Takahashi S.;
 RT "Genetic Polymorphisms of Group B Streptococcus agalactiae scpb."
 RL Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF189003; AAF04281.2; -
 DR EMBL; AF189002; AAF04280.2; -
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR00209; Peptidase_S8.
 DR Pfam; PF02225; PA.1.
 DR Pfam; PF00082; Peptidase_S8; 3.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 FT NON_TER 1130 1130
 SQ SEQUENCE 1130 AA; 124043 MW; E04D343FBA12E994 CRC64;

Query Match 7.5%; Score 257.5; DB 2; Length 1130;
 Best Local Similarity 23.5%; Pred. No. 1.5e-05;
 Matches 118; Conservative 75; Mismatches 167; Indels 143; Gaps 25;

Qy 188 YLKSINAPFGNPDGRGVINIDTGYRHKAMRIDDDAKASMKKEDDKTKKNY-- 245
 Db 110 HVKTLQEKAGK--GAGTVAVVIDAGFKNHEAMRLDTKARYQ-SKENLEKAKKEHGI 165
 Qy 246 ---WLSDKIPAHFYNGKITEVEKYDDGRDYFDPGHMIGIAGNTEODIKNFNGI 301
 Db 166 TYGEWVNDKVAIYHDSKDGKNAVDQ-----EGHTVSG--NAPSEMKKEPYRL 211
 Qy 302 DGIAPNAQIFSYKM---YSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLV 356
 Db 212 EGAMPEAOQLLMRVEIVNGLADYARNYA-----QAIRDAVNLGAKVIMNSFGNALAYA 265
 Qy 357 G--EKYWAIRALKRAGIPMVVATGNVATSASSSSMDLVANNHLMKTGTGNTRTAAHED 414
 Db 266 NLPDETAKAFDYAKSGVSIYTSAGNDSFGKPRPLPLA--DH--PDYGVVGTTPAAADS 320
 Qy 415 AIAVA--SAKNQTFEEDKNIGCESFKRYRNIGAFEDSKITTNEDGTAPSK-LKFEVYIG 471
 Db 321 TLTVASYSPOQLTETATVKTDDHDK-----EMPYISTNR---PEPKAIYAYAN 369
 Qy 472 KGQDODLIGDLRGKIAVMDRIYTRDLKNAFKKAMDKGARAIMVNTVYVYNRD----- 525
 Db 370 RGTKEDDF-KDVEGKIALIER-GDIDFKDKIANAKKAGAVGLI-----YNDQKGFPIE 422
 Qy 526 ---NMTLEPAMGYEADGKTSQVFSISGDGVLKLMNMINPDKITEVKRNKKEDFKKLEQY 583
 Db 423 LPNDQMPAA-----FTSRDGLL-----KD----- 444
 Qy 584 YPIDMESFNKPNVGEKEIDFKFAPDDELYKEDIIVPAGS-----TSMGPRIDL 636
 Db 445 -----NPP-----KITTFNATPK-----VLPTASGTKLSRFSWGTLTADG 479
 Qy 637 LKPDVSAPEGKNIKSTLVINGK 659

Db 480 NIKPDIAAPGODILS--SVANNK 500

RESULT 14

039547

PRELIMINARY; PRT; 731 AA.

AC 039547; 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE PREPRO-CUCUMISIN PRECURSOR.
 OS Cucumis melo (Muskmelon).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
 OC NCBI_TaxID=3656;

SEQUENCE FROM N.A.

STRAIN-TERESA; TISSUE-FRUIT;
 MEDLINE=95105149; PubMed=7806492;
 Yamagata H., Masuzawa T., Nagaoka Y., Ohnishi T., Iwasaki T.;
 "Cucumis, a serine protease from melon fruits, shares structural
 homology with subtilisin and is generated from a large precursor";
 J. Biol. Chem. 269:32725-32731(1994).
 DR EMBL: D32206; BAA06905.1; .
 DR HSSP: P29599; IST3.
 DR MEROPS: S08_092; .
 DR Mendel: 7795; Cucme:1086;7795.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
 DR Signal: Serine protease.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 731 PRE-PRO-CUCUMISIN.
 FT CHAIN 111 731 NATIVE CUCUMISIN.
 FT CHAIN 111 615 MATURE AUTOLYZED CUCUMISIN.
 SO SEQUENCE 731 AA; 78819 MW; 93270A9433A366ED CRC64;

Query Match 7.4%; Score 256; DB 10; Length 731;

Best Local Similarity 24.1%; Pred. No. 9.8e-06; Mismatches 211; Indels 162; Gaps 31;

Matches 148; Conservative 94; Mismatches 211; Indels 162; Gaps 31;

94 NTEPKKEDVYVIA-EFKKESG---EKAIKEL---SLKNTFVLYTDRIENGSAIET 146

25 DSDDGKNITYITMGRKLEPDSALHNRAMLEQVVGSTAPESVLTHTYKRSFNGFAVKL 84

147 TPNDLRIKQIEGISYERAKQVOPMNMHARKEIGVEAIDYLKSNAPFGKNFGRGV 206

85 TEEBAEKIAMEGVSV-----FLNE-INELHTTRSMDFL-----GFPVTPRRSQV 130

207 ISNIDTG---TDYRHKMRIDDOKAKSMRFKEDLCTDKNYWLSDKIPHAENYNGSKI 263

131 ESNLVVGLTGTIMPESPDDEGFSPPRKMGCTCTSNFRCKRIIGARSYHIRPI 190

264 TYEKYDGRDYFDPHGHINGLAGNTEODIKNFNCID-----GIAPNOIEFYKMYSD 318

191 SPGDVNPGRD-TNGHGHNTASTAGGLVSG--ANLYGLGIGTAGRGVPLAKIAIYKCMW 247

319 AGSGFAGDETFMFAIEDSIKHNVDVSVSSG-----FTGTGLVGEKYQAIRALRRAG 371

248 DG---CGDITLAAVDADIDGVDIISLVSGGANPRHYFDALIGS--FHAV-----ERG 298

372 IPMYVATG---NATASASSSDVLANNHLKMTDGTGNVTRTAHEDAIYASAKNOTVE 427

299 ILTNSAGNGCPNFTTASLSPW-----LTSVAATMDRK 334

428 FDKVNIQ-GESFKNRTGAF-----FDKSKITTNEDGTAKPGLKFEVY 469

335 VTQVQIGNGSGFQGVINTFDNQYPLVSGRDIPNTGFDKSTSFCTDKSVNPLLK--- 391

QY 470 IKGQDODLIGDLRGKIAVMDRIYTKOLNARKKAMDKARAIWVNYNYNRDNWTE 529

Db 392 -----GKIYVCEASFG---PHEFFKSLDGAAGVLAQMTSNTDY--AASY-P 430

QY 530 LPAMGYEAD--GTKSGVFISGDDGVKLMNMINPDKTEVKRNKNEEDFDKLEQYYPID 587

Db 431 LPSSVDPNDLTLRTYISIR-SPGATIR-----KSTTI-----LNASAPV- 471

QY 588 MESFNSNKPWVGDEKEIDFKAPDTOKELYEDIIYPAGSTSGPRIDLK-PDVSAPG 646

Db 472 VSEFSSKGPV-----RATKDVIKPDIS-----GPGVEILAMPVSAVPV 510

QY 647 KNIKSTL-NVINGKS 660

Db 511 GIRNRTLFNIISGTS 525

RESULT 15

09FI12

PRELIMINARY; PRT; 840 AA.

AC 09FI12; 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE SERINE PROTEINASE.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;

SEQUENCE FROM N.A.

STRAIN-COLUMBIA;
 RC MEDLINE=99397451; PubMed=10470850;
 Miyajima N., Tabata S.;
 "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
 Sequence features of the regions of 1,011,550 bp covered by seventeen
 P1 and TAC clones";
 RL DNA Res. 6:183-195(1999).
 DR EMBL: AB017065; BAB09160.1; .
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 2.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
 SO SEQUENCE 840 AA; 90616 MW; 0A97E5641FEDDE84 CRC64;

Query Match 7.3%; Score 252.5; DB 10; Length 840;

Best Local Similarity 22.5%; Pred. No. 1.8e-05; Mismatches 233; Indels 201; Gaps 30;

Matches 150; Conservative 82; Mismatches 233; Indels 201; Gaps 30;

56 VSKNPYIDNNTSNEEKIKE-----ENSNSKG-----DYDSFVNKNTEN 97

Db 41 VTLKQPIVHLFEQDELKHKRSKFTPKLRPNRNSRKHGSKIPSVQSHDFLRTKLK- 99

QY 98 PKREDKVVYIAEFKDKESGKAIKELSLKNTKLYLYDDIIFNGSAIETTPNDLTKQI 157

Db 100 -----GEKYK-----LSTYHLINGFALFINSQAEEKSMR 131

QY 158 EGISYERAKQVOPMNMHARKEIGVEAIDYLKSNAPFGKNFGRGVISNIDTGTDR 217

Db 132 KEVANIYLDVSVTRATYTPQFMGLPGA-WVKE---GGFEIAGEGVIIIGIDTGDIPN 186

QY 218 HKAMRIDDDAKASMRFKKE-----DLKGTDKNYWLSDKIPHANVNYNGKITYEKY 268

Db 187 HPSFN-DNDSKRSYPIPKHSGVCEVTPDPSSGCKNKL-----GARHFAQSAVTGICF 240

QY 269 DDGRDYFDP-----HGHIAGILAGNTEODI---KNFNGIDGAPNAQIFSYKMYSDAG 320

Db 241 NSSEDIASFPDGGSHGHHTASVAAGNIGVPIYVSNHNFVGAASGIAPRAPIISYKALKSF 300

QY 321 SGFAGDETFMFAIEDSIKHNVDVSVSSGFT-----GTGLVGEKYQAIRALRRAGI 374

```

Db 301 GGFAD--VVAIDQAOQDGVILSL--ITPNRKPQGVATFFNPIDMALLSAVKAGIFV 356
Qy 375 VVATGNVA-----TSASSSSW-----DLVANNHLKMTDIGNVTRTAHEDAIYASA 421
Db 357 VQAAAGTGPAPKTMSSFSFPIFTVGASSHDRVYNSLTL--GNNVTIPGMGFAIPTDSG 413
Qy 422 KNOTVEFDKVNIGSESFYRNIGAFEDSKITTTNEDGTRKAPSKLKFVYIGKODODLIGL 481
Db 414 K-----MYKMSAFHALNSTVD-----KDMYVGECDYENFDQ 448
Qy 482 D-LRGKIAVMDRIYTKDL--KNAFKKAMD--KGARAIMVYNTVYVYNRDMWTELPAWGY 535
Db 449 DRVSGKLILCS--YSARFVLGLSTIKOALDYAKNLKSATGVIFYIDPY-----VLGF 497
Qy 536 EADEGTSQVFSISGDDGVKLMNMINP-----DKTEVYKRNKEDFKDKLE 581
Db 498 E-----INPTPMDMPGIIIPSVEDSKTLKLYNSSIQRDVYT 534
Qy 582 QYYPIDMESF-----NSNKPNGVDEKEIDFKFAPD-----TDKELYKEDIYVA 625
Db 535 K-----EIVSFGAVNAIEGGLNANFSNRAPKYVYISARGPDPEDNSFNDADVLKPNLVAP- 589
Qy 626 GSTSWG 631
Db 590 GNSIWG 595

```

Search completed: December 7, 2001, 11:49:32
 Job time: 202 sec